# Echantillonnage compressé le long de trajectoires physiquement plausibles en IRM

Nicolas Chauffert

# UNIVERSITÉ PARIS-SUD

**ÉCOLE DOCTORALE : Sciences et Technologie de l'Information, des Télécommunications et des Systèmes**

Équipe PARIETAL, INRIA Saclay
CEA / DSV / I$^2$BM / NeuroSpin

**DISCIPLINE : Physique**

## THÈSE DE DOCTORAT

soutenue le 28/09/2015

par

**Nicolas CHAUFFERT**

# Échantillonnage compressé le long de trajectoires physiquement plausibles en IRM

| | | |
|---|---|---|
| **Directeur de thèse :** | Philippe CIUCIU | Directeur de Recherche (CEA/NeuroSpin) |
| **Encadrant :** | Pierre WEISS | Chargé de Recherche (ITAV Toulouse) |
| **Composition du jury :** | | |
| *Rapporteurs :* | Laurent JACQUES | Professeur (Université catholique de Louvain) |
| | Gabriel PEYRÉ | Directeur de Recherche (Université Paris-Dauphine) |
| *Examinateurs :* | Ali MOHAMMAD-DJAFARI | Directeur de Recherche (Université Paris-Sud) |
| | Justin HALDAR | Assistant Professor (University of Southern California) |
| | Simon MASNOU | Professeur (Université Lyon 1) |

# Contents

# Remerciements

Mes premiers remerciements s'adressent tout naturellement à mes directeurs de thèse. Tout d'abord, à mon directeur « officiel », Philippe Ciuciu. Merci de m'avoir accueilli à NeuroSpin et initié au monde de la recherche, alors que j'étais encore en M2. Merci de m'avoir inculqué la rigueur et de m'avoir montré à quel point il est possible d'être perfectionniste dans le travail. Merci également pour ta patience et ton aide au quotidien.

Merci à Pierre Weiss de m'avoir guidé dans ce travail malgré la distance. Ton enthousiasme sans limite m'a beaucoup boosté durant ces trois années. Merci pour ces petits séjours à Toulouse qui ont été forts agréables, grâce également à la présence d'Alice et d'Anouk.

Cela dit, je ne regrette pas de t'avoir traité de fou[1] devant des millions de téléspectateurs, il fallait bien que la France le sache !


FIGURE 1: Pierre Weiss au journal de 20h de France 2

Merci aux collègues de NeuroSpin pour leur gentillesse et leur disponibilité. Je suis heureux d'avoir pu passer trois ans dans des conditions de travail exceptionnelles, d'un open-space un peu frai l'hiver mais très convivial, au confort d'un bureau chauffé sous la protection d'Edouard. J'en profite pour remercier tout particulièrement Aina, Michael et Philippe (une fois de plus) pour m'avoir soulagé du poids de la logistique à l'approche de la soutenance ! Merci également à Claire et Paul pour leur accueil dans la ville rose.

Je m'attaque désormais à la partie la plus délicate : remercier famille et amis sans en oublier. Je remercie tout d'abord mes parents et ma sœur qui ont toujours été là pour me pousser à étudier. Je leur dois cette curiosité qui m'a aidé durant ma thèse.

[1] le lecteur intéressé pourra s'en rendre compte en visionnant le reportage http://www.francetvinfo.fr/replay-jt/france-2/20-heures/jt-de-20h-du-mardi-3-fevrier-2015_808731.html.

Je remercie également mes amis et collègues du club d'athlétisme de Palaiseau. Les entrainements du mardi et du jeudi soir ont été d'un grand secours pour me vider la tête quand celle-ci était occupée par une démonstration truffée d'erreurs, ou par un code qui ne faisait pas ce qui était prévu... Merci donc à eux de m'avoir soutenu, à commencer par Laurent et Nicolas (dans l'ordre d'obtention du doctorat) pour m'avoir mis la pression après leurs soutenances parfaites, et à Clément, Vincent, aux Marie(s), à Seb, Denis, Antoine, Fifi, Rachou(ille), Guirec, Julie, Suzanne et Michael, pour avoir enchanté entraînements, stages ou vacances.

Une petite pensée pour mes amis de prépas : Martin – pas de discrimination même si tu n'as pas fait de thèse, Jack – merci de m'avoir prêté des cours d'agreg top niveau, mais je ne comprends pas, ça a moins bien marché pour moi ! - et Eline – tu peux me rappeler la date de ta soutenance ?

Au moment de finir d'écrire et de penser aux festivités, je tiens à rassurer Boubou qu'on va bientôt aller boire une petite bière. Ou deux.

Merci aussi aux copains du lycée, Ludo, Thibault, Xavier et Agathe (ainsi qu'à Marjorie et Caro), que de chemin depuis les années à Nerval !

Merci enfin à Marie, d'avoir été à mes côtés durant ces trois années, et de m'avoir toujours soutenu dans ces moments exigeants. A moi de te renvoyer l'ascenseur cette année !

## For english readers

This manuscript is supposed to be written in French. However, most of this PhD work is based on research papers, which are all written in English. Hence, except the background on magnetic resonance imaging and compressed sensing that were written for the French readership, English readers should look at this mansuscript by sequentially reading:

- The summary of this work on page vii.
- The organization and contributions of this thesis on page 24.
- Chapters 2 to 6.

# Résumé

L'imagerie par résonance magnétique (IRM) est une technique d'imagerie non invasive et non ionisante qui permet d'imager et de discriminer les tissus mous grâce à une bonne sensibilité de contraste issue de la variation de paramètres physiques ($T_1$, $T_2$, densité de protons) spécifique à chaque tissu. Les données sont acquises dans l'espace-$k$, correspondant aux fréquences spatiales de l'image. Des contraintes physiques et matérielles contraignent le mode de fonctionnement des gradients de champ magnétique utilisés pour acquérir les données. Ainsi, ces dernières sont obtenues séquentiellement le long de trajectoires assez régulières (dérivée et dérivée seconde bornées). En conséquence, la durée d'acquisition augmente avec la résolution recherchée de l'image.

Accélérer l'acquisition des données est crucial pour réduire la durée d'examen et ainsi améliorer le confort du sujet, diminuer les coûts, limiter les distorsions dans l'image (e.g., dues au mouvement), ou encore augmenter la résolution temporelle en IRM fonctionnelle. L'échantillonnage compressif permet de sous-échantillonner l'espace-$k$, et de reconstruire une image de bonne qualité en utilisant une hypothèse de parcimonie de l'image dans une base d'ondelettes.

Les théories d'échantillonnage compressif s'adaptent mal à l'IRM, même si certaines heuristiques ont permis d'obtenir des résultats prometteurs. Les problèmes rencontrés en IRM pour l'application de cette théorie sont i) d'une part, les bases d'acquisition (Fourier) et de représentation (ondelettes) sont cohérentes ; et ii) les schémas actuellement couverts par la théorie sont composés de mesures isolées, incompatibles avec l'échantillonnage continu le long de segments ou de courbes.

Cette thèse vise à développer une théorie de l'échantillonnage compressif applicable à l'IRM et à d'autres modalités. D'une part, nous proposons une théorie d'échantillonnage à densité variable pour répondre au premier point. Les échantillons les plus informatifs ont une probabilité plus élevée d'être mesurés. D'autre part, nous proposons des schémas et concevons des trajectoires qui vérifient les contraintes d'acquisition tout en parcourant l'espace-$k$ avec la densité prescrite dans la théorie de l'échantillonnage à densité variable.

Ce second point étant complexe, il est abordé par une séquence de contributions indépendantes. D'abord, nous proposons des schémas d'échantillonnage à densité variables le long de courbes continues (marche aléatoire, voyageur de commerce). Ensuite, nous proposons un algorithme de projection sur l'espace des contraintes qui renvoie la courbe physiquement plausible la plus proche d'une courbe donnée (e.g., une solution du voyageur de commerce). Nous donnons enfin un algorithme de projection sur des espaces de mesures qui permet de trouver la projection d'une distribution quelconque sur l'espace des mesures porté par les courbes admissibles. Ainsi, la courbe obtenue est

physiquement admissible et réalise un échantillonnage à densité variable. Les résultats de reconstruction obtenus en simulation à partir de cette méthode dépassent ceux associées aux trajectoires d'acquisition utilisées classiquement (spirale, radiale) de plusieurs décibels (de l'ordre de 3 dB) et permettent d'envisager une implémentation prochaine à 7 Tesla notamment dans le contexte de l'imagerie anatomique haute résolution.


## Abstract

Magnetic Resonance Imaging (MRI) is a non-invasive and non-ionizing imaging technique that provides images of body tissues, using the contrast sensitivity coming from the magnetic parameters ($T_1$, $T_2$ and proton density). Data are acquired in the $k$-space, corresponding to spatial Fourier frequencies. Because of physical constraints, the displacement in the $k$-space is subject to kinematic constraints. Indeed, magnetic field gradients and their temporal derivative are upper bounded. Hence, the scanning time increases with the image resolution.

Decreasing scanning time is crucial to improve patient comfort, decrease exam costs, limit the image distortions (eg, created by the patient movement), or decrease temporal resolution in functionnal MRI. Reducing scanning time can be addressed by Compressed Sensing (CS) theory. The latter is a technique that guarantees the perfect recovery of an image from undersampled data in $k$-space, by assuming that the image is sparse in a wavelet basis.

Unfortunately, CS theory cannot be directly cast to the MRI setting. The reasons are: i) acquisition (Fourier) and representation (wavelets) bases are coherent and ii) sampling schemes obtained using CS theorems are composed of isolated measurements and cannot be realistically implemented by magnetic field gradients: the sampling is usually performed along continuous or more regular curves. However, heuristic application of CS in MRI has provided promising results.

In this thesis, we aim to develop theoretical tools to apply CS to MRI and other modalities. On the one hand, we propose a variable density sampling theory to answer the first inpediment. The more the sample contains information, the more it is likely to be drawn. On the other hand, we propose sampling schemes and design sampling trajectories that fulfill acquisition constraints, while traversing the $k$-space with the sampling density advocated by the theory.

The second point is complex and is thus addressed step by step. First, we propose continuous sampling schemes based on random walks and on travelling salesman (TSP) problem. Then, we propose a projection algorithm onto the space of constraints that

returns the closest feasible curve of an input curve (eg, a TSP solution). Finally, we provide an algorithm to project a measure onto a set of measures carried by parameterizations. In particular, if this set is the one carried by admissible curves, the algorithm returns a curve which sampling density is close to the measure to project. This designs an admissible variable density sampler. The reconstruction results obtained in simulations using this strategy outperform existing acquisition trajectories (spiral, radial) by about 3 dB. They permit to envision a future implementation on a real 7 T scanner soon, notably in the context of high resolution anatomical imaging.

# Chapter 1

# Introduction

L'Imagerie par Résonance Magnétique (IRM) est une technique d'imagerie non-invasive et non ionisante, qui permet d'imager certaines grandeurs physiques à l'intérieur d'un objet, comme la densité de protons par exemple. Un scanner (Fig. 1.1(a)) est utilisé et permet d'obtenir un signal à deux ou trois dimensions que l'on représente sous forme d'image (Fig. 1.1(b)). Nous commençons par présenter brièvement les principes d'acquisition d'images par IRM dans la section 1.1. Nous présentons ensuite quelques notions de Compressed Sensing (CS) (section 1.2, et montrons comment cette théorie d'échantillonnage peut s'appliquer à l'IRM (section 1.3. Cette partie a pour but de donner les motivations principales de cette thèse, et à en fixer les hypothèses de travail. Enfin nous résumons les contributions qui composent le manuscrit et présentons son organisation (sections 1.4 et 1.5).

(a) (b)

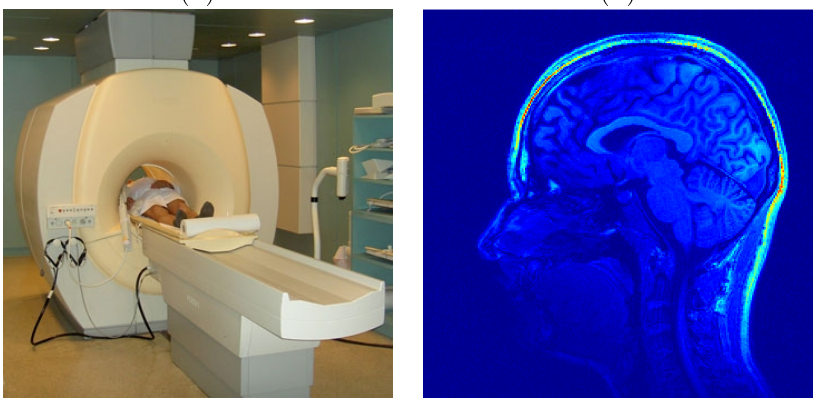

FIGURE 1.1: Illustration d'un scanner IRM (a), et exemple d'image obtenue sur un cerveau humain (b)

## 1.1 L'imagerie par Résonance Magnétique

Nous donnons ici une description très schématique du fonctionnement de l'IRM. Le lecteur intéressé par plus de détails pourra se référer à (Liang and Lauterbur, 2000).

### 1.1.1 Principes de résonance magnétique

Le principe de résonance magnétique a été décrit dans les années 1940 (Bloch, 1946) (Purcell et al., 1946) Il repose sur le fait que certains noyaux (l'hydrogène, le phospore, le sodium) possèdent une propriété quantique appelée spin. Cette grandeur peut être interprétée comme le résultat du mouvement de rotation du noyau autour de son axe. On la note généralement $\vec{\mu}$. En l'absence de champ magnétique extérieur, l'orientation des moments $\vec{\mu}$ est répartie aléatoirement au sein d'un échantillon, et l'aimantation globale est nulle : $\vec{M} = \sum \vec{\mu} = \vec{0}$.

#### Les noyaux en présence d'un champ magnétique

En présence d'un champ magnétique $\vec{\mathbf{B}}_0$, les spins tendent à s'aligner « parallèlement » (faible énergie) ou « anti-parallèlement » (haute énergie) à la direction du champ magnétique. La différence d'énergie $\Delta E$ est proportionnelle à l'intensité du champ magnétique $\vec{\mathbf{B}}_0$:

$$\Delta E = \gamma |\vec{\mathbf{B}}_0| \tag{1.1}$$

où $\gamma$ est le rapport gyromagnétique, et dépend du noyau ($\gamma = 2,675$ $\text{s}^{-1}.\text{T}^{-1}$ pour le noyau d'hydrogène). La statistique de Boltzmann donne le ratio entre le nombre de noyaux à haute énergie $N^-$ et à basse énergie $N^+$:

$$\frac{N^-}{N^+} = e^{-\Delta E / k_B T}$$

où $k_B$ est la constante de Boltzmann ($1,38 \times 10^{-23}$ $\text{J.K}^{-1}$) et $T$ est la température du système. Par exemple, pour le noyau d'hydrogène à température ambiante dans un champ magnétique de 1 Tesla (T), si $N^+ = 1000000$, alors $N^- = 1000006$.

Précisément, les spins ne sont exactement alignés à $\vec{\mathbf{B}}_0$, mais réalisent un mouvement de précession à la fréquence $\omega_0 = \gamma |\vec{\mathbf{B}}_0|$ (fréquence de Larmor) autour de ce vecteur (Fig. 1.2)(b). Les phases dans ce mouvement étant aléatoires, la magnétisation globale $\vec{M}_0$ au sein d'un échantillon est parallèle à $\vec{\mathbf{B}}_0$, comme l'illustre la Fig. 1.2(c).

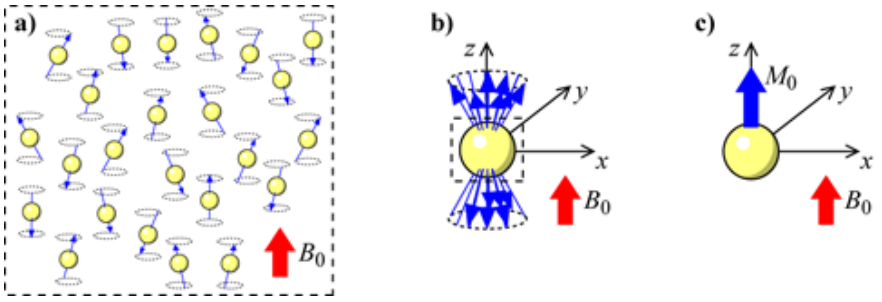


FIGURE 1.2: Illustration du moment magnétique dans un échantillon (a). Les spins précessent autour de l'axe de $\vec{\mathbf{B}}_0$ avec une phase aléatoire (b). L'aimantation macroscopique est dirigée selon $\vec{\mathbf{B}}_0$. (Image tirée de (Poole, 2007)).

La magnétisation $\vec{M}_0$ étant négligeable devant $\vec{\mathbf{B}}_0$, il faut « basculer » ce moment magnétique dans le plan transversal pour pouvoir mesurer $\vec{M}_0$. Une impulsion radio-fréquence est alors utilisée.

### Les impulsions radio-fréquence

Nous considérons désormais un modèle macroscopique, où l'aimantation macroscopique $\vec{M}$ suit un mouvement de précession à la pulsation $\omega_0$ autour de $\vec{\mathbf{B}}_0$. L'application d'un champ magnétique $\vec{\mathbf{B}}_1$ oscillant à la pulsation $\omega_r = \omega_0$ dans le plan transversal provoque un phénomène de résonance et bascule l'aimantation $\vec{M}$ dans le plan transversal suivant une trajectoire d'hélice (Fig. 1.3). A 1 T, pour le noyau d'hydrogène, cela représente une fréquence de 42,58 MHz, ce qui correspond à une onde radio-fréquence (RF): $\vec{\mathbf{B}}_1$ est ainsi également impulsion radio-fréquence.

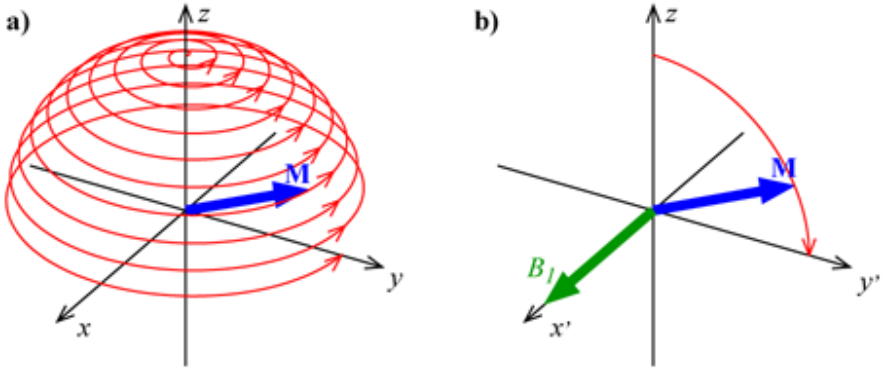


FIGURE 1.3: Évolution de la magnétisation macroscopique $\vec{M}$ sous l'application d'une d'une onde électro-magnétique à la fréquence de résonance. (a) Évolution de l'aimantation dans un repère fixe. (b) Évolution de l'aimantation dans un repère tournant avec $\vec{\mathbf{B}}_1$. Image tirée de (Poole, 2007).

L'effet de l'impulsion radio-fréquence est double : d'une part l'aimantation longitudinale $M_z$ est diminuée à cause du phénomène de bascule. D'un point de vue quantique, l'apport

d'énergie par l'onde fait passer des spins de basse à haute énergie. D'autre part, dans le plan orthogonal, l'aimantation $M_{xy}$ augmente car les spins précessent autour de $\vec{\mathbf{B}}_1$ : il se crée alors un moment magnétique résultant dans cette direction.

### Le phénomène de relaxation

Lorsque l'impulsion radio-fréquence $\vec{\mathbf{B}}_1$ s'arrête, l'aimantation macroscopique $\vec{M}$ retourne dans sa position d'équilibre en suivant une trajectoire hélicoïdale. Le retour à l'équilibre se fait via deux phénomènes indépendants:

- *La relaxation longitudinale.* Les spins retournent dans leur état d'équilibre : cela provoque la repousse de l'aimantation longitudinale $M_z$. La vitesse de repousse dépend du paramètre $T_1$ qui traduit l'interaction des spins avec leur environnement. Précisément:

  $$M_z = M_0\,(1 - \exp(-t/T_1))$$

  où $T_1$ est le temps nécessaire à la récupération de 63% de l'aimantation longitudinale.

- *La relaxation transversale.* En l'absence du champ radio-fréquence, les spins vont à nouveau se déphaser. Le moment résultant dans le plan orthogonal va alors diminuer : c'est la relaxation spin-spin. Précisément :

  $$M_{xy}(t) = M_{xy}(0)\exp(-t/T_2).$$

  Le temps $T_2$ représente le temps pour que la magnétisation transverse retombe à 37% de la magnétisation initiale. En pratique, à cause des inhomogénéités du champ $B_0$, la vitesse de décroissance est plus rapide et régie par le paramètre $T_2^*$. La relation entre $T_2$ et $T_2^*$ est : $\frac{1}{T_2^*} = \frac{1}{T_2} + \frac{1}{T_2'}$ où $\frac{1}{T_2'} = \gamma\Delta B_0$, où $\Delta B_0$ est la différence d'intensité locale de l'intensité du champ $\vec{\mathbf{B}}_0$.

À ce stade, nous n'avons considéré que des grandeurs moyennes au sein d'un échantillon. L'objectif de l'imagerie par résonance magnétique est d'effectuer des cartes de ces grandeurs. En effet, les différents paramètres ($T_1$, $T_2$) sont spécifiques des tissus et permettent une représentation de l'objet comme dans la Fig. 1.1(b). Des grandeurs typiques sont fournies dans le tableau 1.1.

TABLE 1.1: Exemples de valeurs de paramètres $T_1$ et $T_2$ en ms dans le cerveau à 1,5 et 3 Tesla. Étude tirée de (Stanisz et al., 2005).

| | matière blanche | matière grise |
|---|---|---|
| $T_1$ (3 T) | 1100 | 1500 |
| $T_2$ (3 T) | 60 | 70 |
| $T_1$ (1,5 T) | 800 | 1100 |
| $T_2$ (1,5 T) | 80 | 95 |

**Le signal IRM**

Le champ magnétique $M_{xy}$ est mesuré par une antenne réceptrice. Puisque le vecteur de magnétisation oscille autour de l'axe de $\vec{\mathbf{B}}_0$ à la pulsation $\omega_0$ et décroit à la vitesse $T_2^*$ à cause de la relaxation transversale, le signal $S(t)$ a la forme présentée Fig. 1.4.

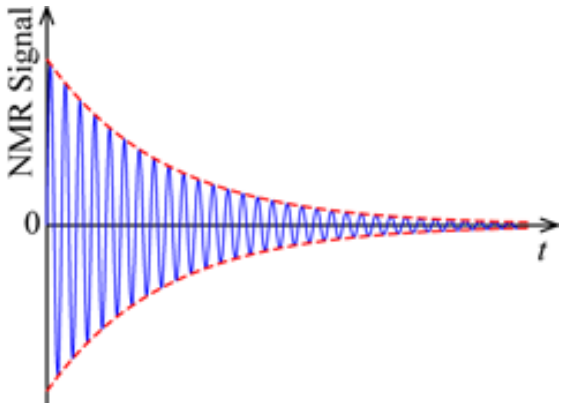


FIGURE 1.4: Signal $S(t)$ avec l'enveloppe de la décroissance exponentielle en $T_2^*$. Image tirée de (Poole, 2007).

### 1.1.2 Utilisation en imagerie

Pour réaliser des images à partir des propriétés magnétiques de la matière, il faut localiser spatialement le signal de résonance magnétique. Pour cela, des gradients de champ magnétique sont utilisés. Il s'agit d'un champ magnétique supplémentaire, orienté selon l'axe de $\vec{\mathbf{B}}_0$ et d'intensité $G(\vec{r}) = (x \cdot G_x, y \cdot G_y, z \cdot G_z)$ où $\vec{r} = (x, y, z)$ (Fig. 1.5). Les quantités $G_x$, $G_y$ et $G_z$ sont des grandeurs scalaires qui peuvent varier dans le temps.

Plaçons le temps $t = 0$ à la fin de l'impulsion radio-fréquence. Le moment magnétique $M_{xy}$ est alors non nul et oscille autour de l'axe de $\vec{\mathbf{B}}_0$ à la pulsation $\omega_0$. En ajoutant un gradient additionnel, la pulsation du mouvement autour de $\vec{\mathbf{B}}_0$ est $\omega(\vec{r}, t) = \omega_0 + \gamma\vec{G}(t).\vec{r}$ et dépend donc de la position du point dans l'objet. Pour simplifier, négligeons l'effet

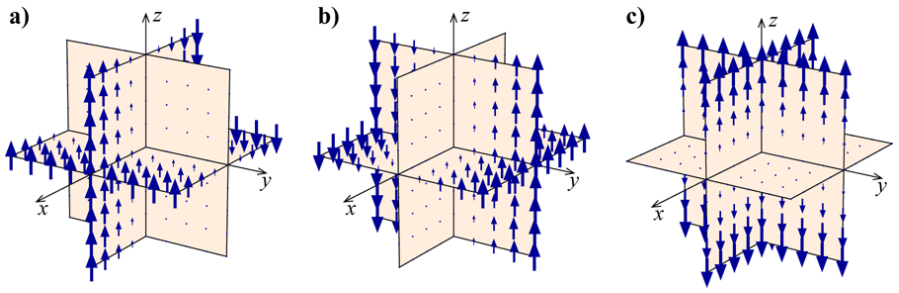


FIGURE 1.5: Illustration des gradients de champ magnétique. a) $G_x$, b) $G_y$, c) $G_z$. (Image tirée de (Poole, 2007)).

de relaxation du signal. Le signal $\mathrm{d}S$ émis par un élément de volume $\mathrm{d}\mathcal{V}$ en $\vec{r}$ est donné par:

$$\mathrm{d}S(\vec{r},t) \propto \rho(\vec{r}) \exp\left(i \int_0^t \omega(\vec{r},t')\mathrm{d}t'\right) \mathrm{d}\mathcal{V},$$

où $\rho(\vec{r})$ est la densité de noyaux en $\vec{r}$. L'antenne réceptrice mesure la somme du signal sur tout le volume, c'est à dire:

$$S(t) \propto \int \rho(\vec{r}) \exp\left(i \int_0^t \omega(\vec{r},t')\mathrm{d}t'\right) \mathrm{d}x\mathrm{d}y\mathrm{d}z$$

Après démodulation par $\omega_0$, le signal obtenu $S'$ est:

$$S'(t) \propto \int \rho(\vec{r}) \exp\left(i\gamma \int_0^t \vec{r}.\vec{G}(t')\mathrm{d}t'\right) \mathrm{d}x\mathrm{d}y\mathrm{d}z$$

Soit maintenant $\vec{k}(t) = -\gamma \int_0^t \vec{G}(t')\mathrm{d}t'$. Alors l'équation précédente s'écrit:

$$S'(t) \propto \int \rho(\vec{r}) \exp\left(-i\,\vec{r}.\vec{k}(t)\right) \mathrm{d}x\mathrm{d}y\mathrm{d}z$$

Cette équation relie le signal à la transformée de Fourier de l'objet $\rho(\vec{r})$ à la fréquence spatiale $\vec{k}(t)$. Cette relation est centrale et indique que les acquisitions en imagerie par résonance magnétique se font dans l'espace de Fourier des fréquences spatiales (appelé ici espace-$k$).

### 1.1.3 Exemple de trajectoires d'acquisition en IRM

#### La sélection de coupe

Les gradients présentés ci-dessus sont appliqués dès la fin de l'impulsion radio-fréquence. Une technique classique consiste à appliquer un gradient dans une direction *pendant*

l'impulsion RF. Cela permet de faire dépendre linéairement la fréquence de résonance dans une direction, et donc d'exciter uniquement une coupe de noyaux dans l'objet. Ainsi, une acquisition IRM peut être réalisée en 2D ou en 3D.

### Exemple de séquences

On appelle répétition l'ensemble des données collectées après une impulsion RF. En théorie, une image IRM peut être acquise en mesurant l'espace-$k$ avec une répétition. Ce type de séquence est utilisée principalement en imagerie fonctionnelle (IRMf) où la résolution temporelle est très importante. Cependant, cela ne permet pas d'obtenir des images de bonne résolution spatiale et crée des artefacts. En effet, le signal décroît assez rapidement (cf Tab 1.1), et une trajectoire trop longue est sensible à cette relaxation. De plus, la vitesse de déplacement dans l'espace-$k$ est limitée par les contraintes de la machine sur la valeur du gradient ($G_{\max}$) et sa dérivée (limitée par le slew-rate $S_{\max}$). Ces deux éléments limitent le nombre de mesures par répétition. En IRM, il est donc commun d'avoir recours à plusieurs répétitions. Une technique d'acquisition très classique consiste à acquérir l'espace-$k$ ligne par ligne (Fig. 1.6).

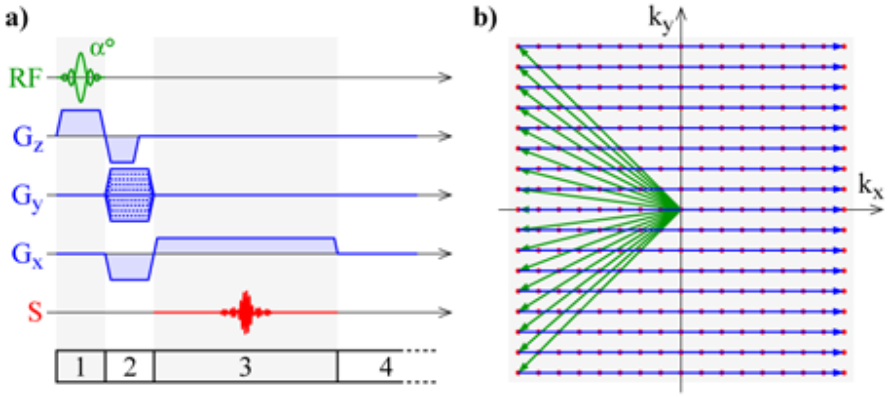


FIGURE 1.6: Chronogramme de séquence d'acquisition de l'espace-$k$ ligne par ligne (a) et parcours correspondant dans l'espace-$k$. Phase 1: excitation d'une seule coupe à l'aide du gradient $G_z$ et l'impulsion RF. Phase 2: déplacement sur le bord gauche de l'espace-$k$ (en vert dans le schéma (b)). Phase 3: Déplacement latéral dans l'espace-$k$ et acquisitions. Le vecteur de gradient est proportionnel à la vitesse de déplacement dans l'espace-$k$. Par exemple, en phase 3, $G_x > 0, G_y = 0, G_z = 0$ correspond à un mouvement rectiligne de la gauche vers la droite de l'espace-$k$. (Image tirée de (Poole, 2007)).

Au cours de cette thèse, nous comparons régulièrement nos trajectoires d'acquisition aux trajectoires EPI (Echo-planar imaging) et spirales qui permettent un parcours de l'espace-$k$ au cours d'une seule répétition. Plusieurs répétitions peuvent toutefois etre utilisées (en entrelaçant les trajectoires) pour augmenter la qualité de l'image (meilleure

résolution, meilleur rapport signal-sur-bruit...). Ces trajectoires sont présentées en figures 1.7 et 1.8.

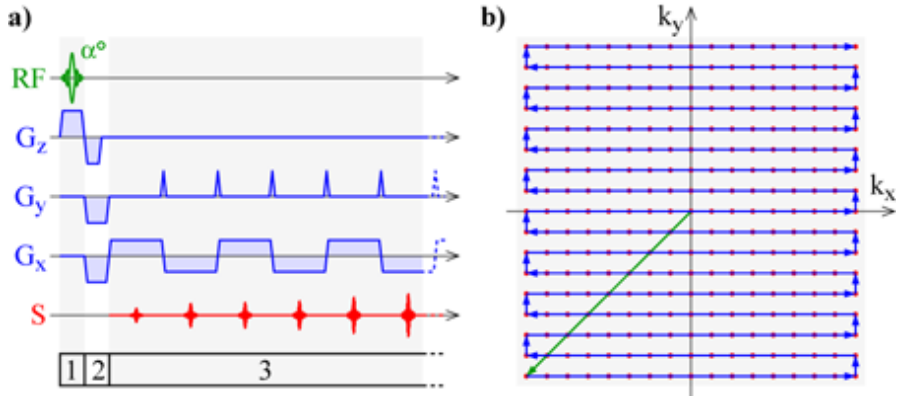


FIGURE 1.7: Chronogramme de la séquence EPI (acquisition de l'espace-$k$ ligne par ligne en une répétition) (a) et parcours correspondant dans l'espace-$k$. Phase 1: excitation d'une seule coupe à l'aide du gradient $G_z$ et l'impulsion RF. Phase 2: déplacement sur un coin de l'espace-$k$ (en vert dans le schéma (b)). Phase 3: Déplacement dans l'espace-$k$ et acquisitions. À nouveau, le gradient $G_x$ permet de se déplacer latéralement dans l'espace-$k$, alors que les décalages verticaux sont réalisées par de petites impulsions (*blips*) selon $G_y$. (Image tirée de (Poole, 2007)).

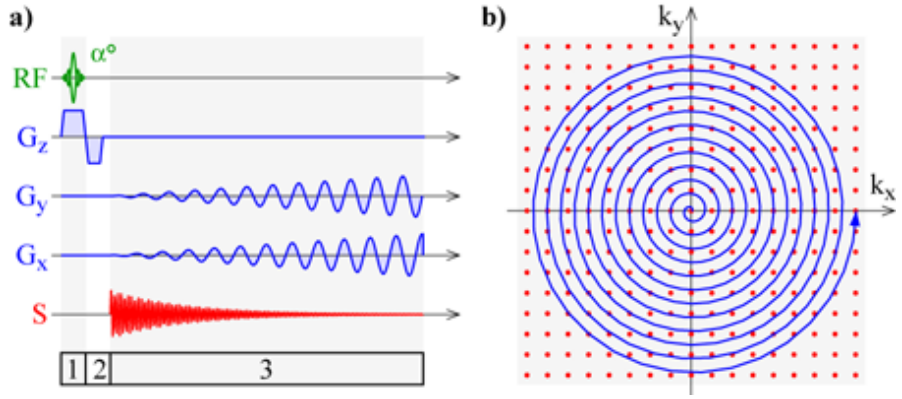


FIGURE 1.8: Chronogramme de la séquence spirale (a) et parcours correspondant dans l'espace-$k$. (Image tirée de (Poole, 2007)).

### Pondération

La présentation de l'imagerie par résonance magnétique faite ici est loin d'être exhaustive. L'objectif est de donner les clés au lecteur pour comprendre les contributions de ce travail. Mentionnons toutefois qu'en IRM, il est possible d'obtenir plusieurs types d'informations, à savoir des cartes dépendant d'un paramètre parmi le $T_1$, $T_2$, $T_2^*$, et la densité de protons. Pour cela, il suffit de faire varier les paramètres d'acquisition TR (le temps entre deux répétitions, variant de 500 à 1500 ms) et TE (temps d'écho: le temps entre l'impulsion RF

et la mesure du signal de quelques ms à 100 ms). Des images avec différentes pondérations sont présentées Fig. 1.9.

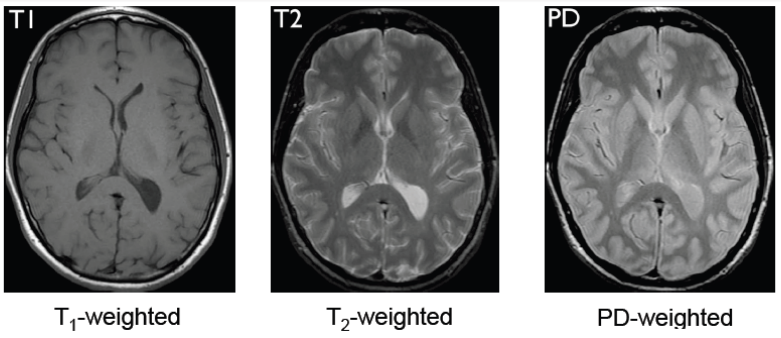


FIGURE 1.9: Exemple d'images IRM avec des pondérations en $T_1$ à gauche (TE et TR courts), $T_2$ au centre (TR et TE longs), et en densité de proton à droite (TR longs et TE courts).

Une **image pondérée** $T_1$ est obtenue en appliquant des TR courts. En effet, d'une excitation à la suivante, le signal disponible sera proportionnel à la repousse longitudinale du signal à la fin de la répétition précédente, et dépendra donc $T_1$. Un TE court évite un trop grand déphasage dans le plan orthogonal à $\vec{\mathbf{B}}_0$ et empêche que le signal dépende de $T_2$.

Rappelons qu'après une excitation RF, les protons perdent la cohérence de leur phase dans leur mouvement de rotation dans le plan orthogonal à $\vec{\mathbf{B}}_0$. Cette perte de cohérence est due à l'interaction spin-spin (décroissance $T_2$) et aux inhomogonéités locales de champ magnétique ($T_2^*$). Une **image pondérée** $T_2^*$ est simplement obtenue en considérant des longs TR (la repousse est complète et ne dépend plus de $T_1$). Pour obtenir une image pondérée $T_2$, on utilise que le déphase dû aux inhomogénéités de $\vec{\mathbf{B}}_0$ est réversible. En effet, soit $t = 0$ le temps à la fin de l'impulsion RF. Après celle-ci, les spins se déphasent car leur vitesse de rotation dépend de la fréquence de Larmor locale. À $t = TE/2$, on applique une impulsion RF deux fois plus longue (la rotation est alors de 180°) qui renverse les spins. Les spins les plus rapides sont désormais en retard sur les spins les plus lents. À $t = TE$, les spins se réalignent et l'effet des inhomogonéités du champ $\vec{\mathbf{B}}_0$ sont annulés. Cette séquence s'appelle *écho de spin* et permet de réaliser une **image pondérée** $T_2$ (en considérant des TR longs pour éviter la dépendance en $T_1$).

Une alternative aux séquences écho de spin sont les séquence à *écho de gradient*. Après l'excitation RF, des gradients de champ magnétique sont appliqués et déphasent le signal. Le signal est alors maximal lorsque les gradient symétriques ont été joués (le déphasage dû aux gradients est alors annulé). Par exemple, dans la Fig. 1.6, pour chaque impulsion RF, l'écho est obtenu au milieu de la phase 3, i.e., au milieu de la ligne acquise, car le gradient déphasage du au gradient de lecture $G_x$ est compensé à ce moment-là (c'est le

temps d'écho d'une séquence écho de gradient). Dans cette thèse, nous considérerons des séquences à écho de gradient, même si nous ne cherchons pas particulièrement à mesurer l'écho, car celui correspondrait à un passage par le centre de l'espace-$k$.

L'avantage des séquences écho de gradients est qu'il est possible de basculer l'aimantation d'un angle $\alpha$ <90°. Le signal disponible ainsi que le rapport signal-sur-bruit sont plus faibles. L'avantage est que la durée d'attente pour la repousse du signal est plus faible. On a donc un compromis entre rapport signal-sur-bruit et TR.

### 1.1.4 Les limites de l'IRM

#### Un compromis résolution spatiale-résolution temporelle

Le nombre de données nécessaire à collecter dans l'espace-$k$ est lié à la résolution à laquelle on souhaite reconstruire l'image : c'est le théorème de Shannon (Mallat, 1999). Cependant, la collecte des données prend du temps car la vitesse de déplacement dans l'espace-$k$ est limitée par les paramètres des gradients. Accélérer l'acquisition en IRM est un challenge crucial car cela permet d'augmenter le confort du patient (environnement bruyant et exigu), et peut permettre de diminuer les coûts dans un contexte clinique. De plus, cela peut permettre d'augmenter la résolution temporelle en imagerie fonctionnelle En IRMf de repos, réduire le TR permet de réduire les effets du mouvement du patient, ou d'éviter les artefacts de repliement (TR $<$ 1 s). En IRMf d'activation (i.e., le sujet doit effectuer une tâche pendant l'acquisition), réduire le TR permet de mieux estimer la fonction de réponse hémodynamique, qui correspond à la réponse impulsionnelle du couplage neuro-vasculaire (Ciuciu et al., 2003). Une meilleure estimation de cette fonction permet de recouvrer des activations plus fines dans certaines régions cérébrales impliquées dans les tâches à réaliser par le sujet au cours d'un paradigme (Handwerker et al., 2004; Badillo et al., 2013). A temps d'acquisition fixé, la problématique réciproque est d'améliorer la qualité et la résolution de l'image.

A très haut champ[1] ($|\vec{\mathbf{B}}_0| \geq 7$T), le rapport signal-sur-bruit (SNR) est accru dans les images (le signal disponible est plus fort, et on a la formule $SNR \propto |\vec{\mathbf{B}}_0|$). Il est donc possible d'augmenter la résolution spatiale. Ainsi l'ensemble des fréquences à mesurer dans l'espace-$k$ est plus conséquent, et fait exploser le temps d'acquisition. Par exemple l'acquisition de données d'une partie de cerveau à la résolution $300\mu m$ isotrope (image de $640 \times 640 \times 192$ pixels) a nécessité 1 heure et 20 minutes d'acquisition.

[1] Le centre d'acquisition NeuroSpin possède un scanner clinique de 7 T et attend un scanner de 11.7 T pour février 2016: une première mondiale !

**Méthodes existantes pour accélérer l'acquisition.**

L'imagerie parallèle (IRMp) a été la première méthode proposée pour accélérer l'acquisition IRM (Sodickson and Manning, 1997; Pruessmann et al., 1999; Griswold et al., 2002) et consiste à utiliser plusieurs antennes pour la réception du signal. Ainsi chacune mesure une partie de l'image, qui est finalement reconstruite par une reconstitution de l'espace-$k$ (méthode GRAPPA) (Sodickson and Manning, 1997; Griswold et al., 2002), ou par une reconstruction directe dans le domaine image (méthode SENSE) (Pruessmann et al., 1999). L'utilisation de la redondance d'information collectée permet de réduire le nombre de mesures et donc d'accélérer les acquisitions. D'autres techniques proposent d'acquérir simultanément plusieurs coupes en IRMf avec une séquence EPI (Feinberg et al., 2010), et permettent d'acquérir des volumes d'image en moins d'une seconde contre 2 à 4 en imagerie fonctionnelle classique.

Dans cette thèse, nous développons des méthodes qui permettent d'accélérer l'acquisition d'images IRM en échantillonnant l'espace-$k$, et en utilisant des informations *a priori* sur l'image à reconstruire. Les méthodes utilisées dans cette thèse se basent sur la théorie du Compressed Sensing (ou échantillonnage compressif) introduit dans la communauté de l'IRM par (Lustig et al., 2008). Ces techniques peuvent être combinées avec l'IRMp. Cependant, dans ce travail, nous nous sommes limités à l'imagerie « classique », i.e, avec une seule antenne. L'impact de l'utilisation de plusieurs antennes sur les stratégies d'échantillonnage est illustré par exemple dans (Florescu et al., 2014).

## 1.2 Introduction à la théorie de l'échantillonnage compressif

Dans cette thèse, nous traiterons des images discrètes, c'est à dire définies sur une grille finie de pixels. Nous les noterons souvent sous forme d'un vecteur de $\mathbb{R}^n$ où $n$ est le nombre de pixels. En imagerie par résonance magnétique, comme dans beaucoup d'autres d'autres systèmes, le signal est observé via un ensemble de mesures linéaires (e.g., des coefficients de Fourier en IRM). Ainsi, il existe un vecteur $y \in \mathbb{R}^m$ et une matrice $A \in \mathbb{R}^{m \times n}$ tels que l'ensemble des mesures s'écrive:

$$y = Ax$$

Si la matrice $A$ est inversible (en particulier $m = n$), la reconstruction du signal se fait en inversant la matrice $A$. Par exemple, en imagerie par résonance magnétique, lorsque les mesures se font sur une grille Cartésienne, l'image s'obtient par transformée de Fourier inverse. L'échantillonnage compressif s'intéresse au cas où $m \ll n$, ce que nous supposerons désormais.

### 1.2.1 Parcimonie et incohérence : les clés du Compressed Sensing

Lorsque $m < n$, l'ensemble des solutions de l'équation $y = Ax$ est infini. Pour pouvoir retrouver $x$ à partir de $y$, il faut donc avoir d'autres informations sur le vecteur $x$. Une hypothèse généralement formulée est que $x$ est parcimonieux (resp. compressible), c'est à dire qu'il peut être représenté (resp. approché) par un vecteur de $\mathbb{R}^n$ contenant $s$ coefficients non nuls, avec $s \ll n$. C'est le cas notamment des signaux sonores via la représentation de Fourier à fenêtre, ou des images dans des bases d'ondelettes (le format de compression JPEG-2000 utilise une telle représentation). Ainsi, quitte à modifier la matrice $A$, on suppose que $x$ est parcimonieux. Une formulation naturelle pour retrouver x est le problème:

$$\arg\min_{y=Az} |z|_0$$

où $|z|_0$ est la pseudo-norme qui compte le nombre de coefficients non-nuls de $z$. Cependant, ce problème est NP-complet, et ne peut être résolu pour des problèmes de taille intéressante. On considère généralement le problème convexe suivant, qui permet d'obtenir des signaux parcimonieux également :

$$\arg\min_{y=Az} \|z\|_1 \tag{1.2}$$

où $\|z\|_1 = \sum_{i=1}^{n} |z_i|$.

Les premiers résultats de Compressed Sensing (Candès et al., 2006a; Candès, 2008; Donoho, 2006), donnent des garanties de reconstructions reposant sur la propriété RIP : *Restricted Isometry Property*. Une matrice $A$ vérifie la condition RIP s'il existe $\delta_s \geqslant 0$ tel que, quelque soit $x$ $s$-parcimonieux :

$$(1-\delta)\|x\|_2 \leqslant \|Ax\|_2 \leqslant (1+\delta)\|x\|_2. \tag{1.3}$$

Cette condition permet d'obtenir le résultat suivant (Candès, 2008) :

**Theorem 1.1.** *Si $\delta_{2s} \leqslant \sqrt{2}-1$, alors $x$ est l'unique solution de (1.2).*

Ce résultat est uniforme au sens où si cette propriété est vérifiée pour $A$, alors *tous* les vecteurs $s$-parcimonieux pourront être reconstruits exactement. Les inconvénients de cette approche sont multiples. À part pour des matrices très spécifiques (matrices aléatoires à entrées indépendentes identiquement distribuées (i.i.d). par exemple) ne correspondant pas à des systèmes physiques, la constante RIP ne peut pas être calculée. De plus, la condition RIP est très restrictive car elle garantit la reconstruction de *tous*

les signaux $s$-parcimonieux. Or, dans les applications, l'objectif est de reconstruire un seul signal. On peut donc espérer obtenir des conditions de reconstruction exacte bien moins restrictives que le RIP.

Des résultats de reconstruction non uniformes (i.e., valables pour un signal $x$ donné) sont apparus plus récemment (Rauhut, 2010; Candès and Plan, 2011). Considérons une matrice orthogonale

$$A_0 = \begin{pmatrix} a_1^* \\ \vdots \\ a_n^* \end{pmatrix}$$

et construisons $A$ de la façon suivante :

$$A = \begin{pmatrix} a_{J_1}^* \\ \vdots \\ a_{J_n}^* \end{pmatrix},$$

où les entiers $\{J_1, \dots, J_n\}$ sont tirés de manière i.i.d. et uniformément dans l'ensemble $\{1, \dots n\}$. Alors, un résultat typique s'écrit :

**Theorem 1.2.** *Soit $x$ un vecteur $s$-parcimonieux. Si le nombre de mesures $m$ vérifie:*

$$m \geqslant C \cdot s \cdot \left( n \cdot \max_{1 \leqslant k \leqslant n} \|a_k\|_\infty^2 \right) \cdot \log\left(\frac{n}{\eta}\right) \tag{1.4}$$

*où $C$ est une constante universelle, alors $x$ est l'unique minimiseur de (1.2) avec probabilité $1 - \eta$.*

Ce théorème est riche en enseignements car il fait intervenir deux grandeurs cruciales de la théorie de l'échantillonnage compressif. D'une part, le nombre de mesures nécessaires $m$ est proportionnel à la *parcimonie* $s$. Comme expliqué précédemment, sans information *a priori*, il aurait fallu $m = n$ mesures, alors que si $s \ll n$, on peut espérer reconstruire le signal à partir de $m \ll n$ mesures. L'autre grandeur cruciale est la *cohérence* $\kappa(A_0) = n \cdot \max_{1 \leqslant k \leqslant n} \|a_k\|_\infty^2$. La cohérence est comprise entre 1 et $n$ pour des matrices orthogonale. Pour la transformée de Fourier discrète $F$, $\kappa(F) = 1$, ce qui est un cas optimal. Un des pire cas est l'identité car $\kappa(I_n) = n$. La cohérence mesure comment un atome de la base dans laquelle le signal est parcimonieux s'« étale » dans la base d'acquisition. Mieux l'énergie de ces atomes est répartie dans la base d'acquisition, plus une mesure fournit d'information sur le signal.

En imagerie par résonance magnétique, la matrice $A_0$ est égale à $F^*\Psi$ où $\Psi$ est la matrice d'une transformée en ondelettes orthogonale inverse. Dans ces conditions, $\kappa(A_0) = \mathcal{O}(n)$,

ce qui rend les résultats d'échantillonnage compressif non applicables en IRM : c'est la « barrière de la cohérence » (Adcock et al., 2013).

### 1.2.2 Casser la « barrière de la cohérence »

La cohérence $\kappa$ est un critère global sur les bases de représentation et d'acquisition. Beaucoup de matrices ont ainsi une distribution de $\|a_i\|_\infty^2$ très hétérogène, comme c'est le cas de la matrice $F^*\Psi$ en IRM[2]. Un exemple pathologique est introduit dans le papier (Bigot et al., 2013) où les auteurs considèrent la matrice :

$$A_0 = \begin{pmatrix} 1 & 0 \\ 0 & F_{n-1} \end{pmatrix}.$$

Cette matrice vérifie $\kappa(A_0) = 1$. Dans ce cas, le théorème 1.2 n'est pas intéressant car le nombre de mesures doit être $\mathcal{O}(n)$. Cependant, en mesurant la première composante du signal de manière déterministe et les autres échantillons de manière aléatoire, le théorème 1.2 assure la reconstruction dès que $m \geqslant 1 + C \cdot s \cdot \frac{n}{n-1} \cdot \log\left(\frac{n}{\eta}\right)$ avec probabilité $1 - \eta$. Un échantillonnage déterministe des échantillons correspondant aux grandes valeurs de $\|a_i\|_\infty^2$ est une manière simple de casser la barrière de cohérence. Une formalisation de cette technique est donnée par le théorème 2.9 page 36.

Une seconde méthode pour casser la barrière de cohérence est de réaliser un échantillonnage à densité variable : les échantillons les plus cohérents (ce sont les seuls qui contiennent l'information de certains atomes de la bases de représentation) sont plus importants et doivent être tirés avec une plus grande probabilité (Krahmer and Ward, 2014; Puy et al., 2011; Chauffert et al., 2014a). Précisément, le théorème suivant, prouvé dans (Rauhut, 2010; Candès and Plan, 2011) permet de déterminer une distribution « optimale » qui justifie l'échantillonnage à densité variable:

**Theorem 1.3.** *Soit $x$ un vecteur $s$-parcimonieux et $\pi$ la distribution définie par*

$$\pi_k = \frac{\|a_k\|_\infty^2}{\sum_{j=1}^n \|a_j\|_\infty^2}.$$

*Si $A$ est construite en tirant $m$ lignes de $A_0$ suivant $\pi$, et si*

$$m \geqslant C \cdot s \cdot \left(\sum_{j=1}^n \|a_j\|_\infty^2\right) \cdot \log\left(\frac{n}{\eta}\right)$$

[2] voir la figure 2.3(b-c) page 38

*où $C$ est une constante universelle, alors $x$ est l'unique minimiseur de* (1.2) *avec probabilité* $1 - \eta$.

Ce théorème semble plus applicable à l'IRM car on peut montrer $\sum_{j=1}^{n} \|a_j\|_\infty^2 = \mathcal{O}(\log(n))$ (Krahmer and Ward, 2014). Ainsi, le nombre de mesures nécessaires pour reconstruire une image IRM est $\mathcal{O}(s \log(n)^2)$, dans le cas où $A = F^* \Psi$. Cependant, pour une image de taille $256 \times 256$, $\log(n)^2 \simeq 123$, ce qui rend le résultat peu intéressant. En effet, $C \geqslant 1$, et si $s$ est de l'ordre de $0,2 \cdot n$ (ce qui correspond à une bonne approximation de l'image), le nombre de mesures $m$ dépasse $n$...

En pratique, le théorème 1.3 est utilisé pour donner une idée sur la stratégie d'échantillonnage dans le cadre de simulations d'échantillonnage pour l'IRM, et un nombre de mesures de l'ordre de $m = 0,2 \cdot n$ permet une très bonne reconstruction de l'image, même si les conditions requises par le théorème sont loin d'être remplies ! Une réponse à ce phénomène est donnée dans (Adcock et al., 2013). L'hypothèse effectuée jusqu'à présent est la parcimonie du signal IRM. Or le signal possède une structure, sa parcimonie est différente dans chaque sous-bande de sa décomposition en ondelettes (Fig. 1.10).

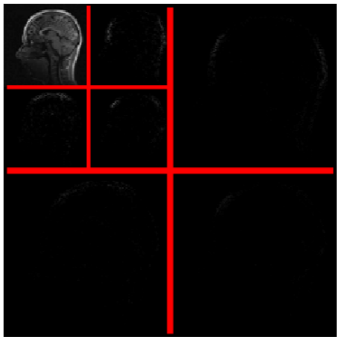

FIGURE 1.10: Décomposition d'une image de cerveau dans une base d'ondelettes.

Les auteurs de (Adcock et al., 2013) mettent en évidence l'importance de la structure de la parcimonie par le *flip test*. Sur des schémas classiques dans le cadre de l'IRM par exemple, leur conclusion est que si les coefficients non nuls sont permutés (le vecteur $x$ est renversé), alors la reconstruction ne fonctionne plus (en particulier, la matrice $A$ en IRM ne vérifie pas la condition RIP).

La parcimonie dans une base d'ondelettes croît avec la résolution (Fig. 1.10). Des résultats faisant intervenir la structure de la parcimonie sont apparus récemment (Boyer et al., 2015a; Adcock et al., 2013). Le nombre de mesures nécessaires est alors plus faible que celui donné par le théorème 1.3, et la distribution d'échantillonnage est légèrement différente. Elle est plus importante au centre, car ce sont les basses fréquences qui contiennent l'essentiel de l'information du signal.

Dans cette thèse, nous nous sommes limités à des tirages selon la distribution fixée par le théorème 1.3, ou même selon des distributions heuristiques. La plupart des méthodes présentées dans cette thèse sont génériques et s'adaptent à n'importe quelle distribution d'échantillonnage.

### 1.2.3 Remarques

- Dans les systèmes physiques, les données observées sont souvent contaminées par du bruit. En imagerie par résonance magnétique, celui-ci provient de l'agitation thermique à l'intérieur de l'objet à imager, ainsi que des appareil électroniques. Celui-ci peut être modélisé en première approximation par un bruit blanc gaussien additif (Aja-Fernández and Tristán-Vega, 2013). L'acquisition peut alors être modélisée par l'équation:

$$y = Ax + n$$

où $n_i \overset{i.i.d.}{\sim} \mathcal{N}(0, \eta^2)$, $1 \leqslant i \leqslant m$. Les résultats de reconstruction présentés précédemment sont stables par rapport au bruit, et également par rapport à l'approximation du signal par un signal $s$-parcimonieux, si celui-ci est seulement compressible. Précisément, la méthode de reconstruction est modifiée en :

$$\arg\min_z \frac{1}{2}\|y - Az\|_2^2 + \lambda\|z\|_1. \tag{1.5}$$

Alors, avec le même nombre de mesures que dans le théorème 1.3, et en prenant $\lambda = 10 \cdot \eta \cdot \log\sqrt{n}$, la reconstruction est stable au sens où :

$$\|x - z^*\| \leqslant C\left(\frac{\|x - x_s\|_1}{\sqrt{s}} + \sqrt{\frac{s\log(n)}{m}}\eta\right),$$

où $z^*$ est la solution de (1.5), et $x_s$ est le vecteur $s$-parcimonieux formé des $s$ plus grandes composantes de $x$ en valeur absolue (Candès and Plan, 2011).

- Dans cette thèse, nous nous limitons à des reconstructions dans des bases d'ondelettes orthogonales. Cependant, en pratique, il est possible d'obtenir des meilleures reconstructions qu'avec le problème 1.5, qui crée des artefacts d'ondelettes. Des méthodes reposant sur des ondelettes redondantes ou utilisant une pénalisation par variation totale ont été proposées (Block et al., 2007; Boyer et al., 2012; Florescu et al., 2014). Des garanties semblables aux théorème 1.3 commencent à apparaître dans le cas de bases redondantes (Poon, 2015b) et pour des régularisations par variation totale (Poon, 2015a).

## 1.3 L'échantillonnage compressif en IRM : principes et limites actuelles.

La théorie de l'échantillonnage compressif a été rapidement introduite dans la communauté IRM (Lustig et al., 2007). Avant de discuter des résultats existants, rappelons les contraintes d'échantillonnage en IRM.

### Les contraintes d'échantillonnage en IRM

Le déplacement dans l'espace-$k$ s'effectue grâce aux gradients de champ magnétique $\vec{G}(t)$ : $\dot{\vec{k}}(t) = -\gamma \cdot \vec{G}(t)$. Ceux-ci étant créés par des courants dans trois bobines (chacune créant un champ magnétique dans une direction), ils sont limités en intensité, et leur dérivée est également bornée:

$$\|G(t)\| \leqslant G_{\max}, \quad \|\dot{G}(t)\| \leqslant S_{\max}, \quad \forall t \in [0, T], \tag{1.6}$$

où $T$ est le temps d'acquisition le long de la trajectoire $k(t)$. La norme utilisée est soit la norme $\|\cdot\|_2$, soit la norme $\|\cdot\|_\infty$, suivant si le courant dans chacune des trois bobines est issu d'un même générateur ou non. Une trajectoire d'acquisition dans l'espace-$k$ doit donc appartenir à l'ensemble suivant :

$$\mathcal{S}_T = \left\{ s : [0, T] \to \mathbb{R}^3, \|\dot{s}(t)\| \leqslant \gamma \cdot G_{\max}, \|\ddot{s}(t)\| \leqslant \gamma \cdot S_{\max}, \forall t \in [0, T] \right\}. \tag{1.7}$$

De plus, les gradients sont appliqués après la fin de l'excitation RF, ce qui impose $s(0) = 0$. Si le temps de répétition est fixé, on peut également supposer $s(k \cdot TR) = 0$, i.e., la trajectoire repart du centre de l'espace-$k$ tous les $TR$. Nous pouvons donc considérer un ensemble de contraintes affines $\mathcal{A}$. Une trajectoire IRM est jouable si elle vérifie les contraintes cinématiques, i.e., $\in \mathcal{S}_T$, et si elle vérifie les contraintes affines, i.e., $\in \mathcal{A}$, dépendantes de la séquence.

### L'application de l'échantillonnage compressif en IRM

La première utilisation de l'échantillonnage compressif en IRM (Lustig et al., 2007) a cherché à adapter directement les résultats du type du théorème 1.2. Les auteurs ont proposé d'utiliser un échantillonnage à densité variable de manière heuristique. Cependant, les schémas obtenus à partir de tirages i.i.d. (Fig. 1.11(a)) ne sont pas faisables en 2D car ils ne sont même pas sur une courbe continue, et n'ont donc aucune chance d'appartenir à $\mathcal{S}_T$. La solution proposée est de parcourir l'espace-$k$ le long de lignes

parallèles dans la troisième dimension (Fig. 1.11(b)), en effectuant plusieurs répétitions. En ce sens, cette stratégie d'acquisition est très proche de celle proposée en Fig. 1.6, s'avère donc réalisable en pratique, et très simple à mettre en œuvre.

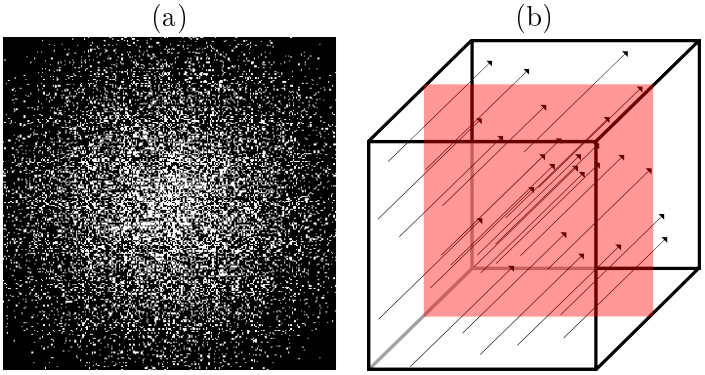


FIGURE 1.11: Schéma d'échantillonnage aléatoire 2D proposé dans (Lustig et al., 2007) (a). Acquisition selon la troisième direction (direction de lecture) (b).

Les trajectoires ainsi obtenues sont physiquement réalisables, mais ne relèvent plus de la théorie de l'échantillonnage compressif. De nouvelles théories sont apparues *a posteriori* (Bigot et al., 2013; Boyer et al., 2015a) pour décrire une théorie de l'échantillonnage par « blocs de mesures » (par exemple, des lignes), mais les lignes parallèles sont loin d'être optimales selon cette théorie.

D'autres méthodes utilisent une pré-modulation du signal pendant l'impulsion RF pour diminuer la cohérence entre les bases d'acquisition (qui n'est plus la transformée de Fourier) et la base de représentation. Cette méthode appelée *spread spectrum* a été introduite dans (Puy et al., 2012a; Haldar et al., 2011). L'objectif de ces tentatives est de se rapprocher de la théorie de l'échantillonnage compressif. L'inconvénient est que la structure du signal est détruite, et en particulier, il n'est plus possible de promouvoir les basses fréquences qui contiennent l'information du signal.

### Trajectoires d'échantillonnage en IRM

Bien avant la théorie de l'échantillonnage compressif, des trajectoires d'acquisitions partielles de l'espace-$k$ sont apparues pour accélérer les acquisitions IRM. Les deux grands types de trajectoires utilisés pour sous-échantillonner l'espace-$k$ en IRM sont les trajectoires spirales (Spielman et al., 1995; Tsai and Nishimura, 2000; Kim et al., 2003; Park et al., 2005) et radiales (Feng et al., 2014). Ces trajectoires sont facile à implémenter et vérifient les contraintes, ce qui explique leur succès. D'autres types de trajectoires sont également rencontrées dans la littérature : des spirales bruitées (Lustig et al., 2007), des trajectoires en rosette (Noll, 1997), des trajectoires 3D hélicoïdales (Shu et al., 2006)...

Ces méthodes n'ont pas de justification théorique mais sont facile à implémenter sur des scanners et permettent d'acquérir des images d'assez bonne qualité. Cependant, les théories de l'échantillonnage laissent penser qu'un sous-échantillonnage plus important de l'espace-$k$, et donc un gain en temps d'acquisition est possible. Les premières justifications théoriques concernant des trajectoires continues (radiales, spirales, et courbes de Hilbert) ont reposé sur un calcul de la cohérence mutuelle (Willett, 2011)[3]. Récemment, une série d'articles (Polak et al., 2012a; Bigot et al., 2013; Boyer et al., 2014) a généralisé les résultats d'échantillonnage compressif pour des ensemble de mesures fixés à l'avance. L'idée est de se donner un ensemble de courbes admissible (dans $\mathcal{S}_T$) et de faire un tirage i.i.d. parmi ces courbes, selon une distribution qui peut être calculée explicitement. Ceci permet de réaliser un échantillonnage physiquement plausible, mais n'est pas générique car il dépend d'un choix *a priori* d'un ensemble de courbes. Toutefois, cette approche généralise les schémas radiaux et spirales, car il est possible d'inclure des segments partant du centre de l'espace-$k$ ainsi que des trajectoires spirales dans l'ensemble des courbes que l'on se donne.

D'autres travaux ont tenté de générer des trajectoires d'échantillonnage dont l'objectif est de mesurer une grille suffisamment fine de l'espace d'acquisition afin de vérifier le critère de Shannon (Unnikrishnan and Vetterli, 2013; Gröchenig et al., 2014) lorsque le signal est à bande limitée.

### Problématique

Les expériences d'échantillonnage compressé en deux dimensions (Knoll et al., 2011; Chauffert et al., 2013b) ont montré que la densité d'échantillonnage est cruciale pour obtenir des bons résultats de reconstruction. L'approche classique qui consiste à effectuer un tirage aléatoire en deux dimension et de mesurer dans la troisième direction comme présenté Fig. 1.11(b) est sous-optimale, car dans la troisième direction, la densité d'échantillonnage est uniforme. Or, en deux dimensions, les meilleures densités d'échantillonnage sont denses au centre, et décroissent lorsque l'on s'éloigne du centre. Nous montrerons qu'en trois dimensions également, les densités à décroissance radiale permettent d'obtenir les meilleurs résultats de reconstruction, et donc que l'approche classique est sous-optimale.

L'objectif de cette thèse est donc le suivant: [modifier]

> Comment, à temps d'acquisition ou à nombre de mesures fixé, réaliser le schéma d'échantillonnage physiquement admissible ?

[3] Le papier contenait des erreurs corrigées dans (Willett., 2011)

## 1.4 Contributions

Les contributions présentées dans cette thèse sont les suivantes:

- Dans le Chapitre 2, nous reformulons les théorèmes d'échantillonnage compressif de (Rauhut, 2010; Candès and Plan, 2011) à la manière du théorème 1.3, de manière à définir une distribution *optimale* dépendant des bases d'acquisition et de représentation (Fig. 1.12(a)). Comme expliqué précédemment, un échantillonnage à densité variable permet de casser la « barrière de la cohérence ». Nous introduisons une seconde méthode pour casser cette barrière reposant sur l'échantillonnage déterministe des basses fréquences de l'espace-$k$ (Théorème 2.9) illustré Fig. 1.12(b). La densité d'échantillonnage optimale est enfin comparée à des distributions heuristiques sur des simulations en 2D.

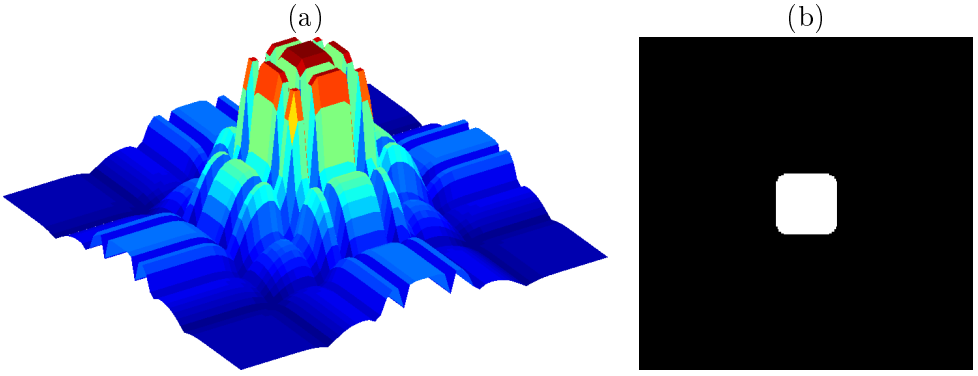


FIGURE 1.12: Distribution $\pi$ *optimale* (a) et schéma d'échantillonnage déterministe des basses fréquences correspondant aux grandes valeurs de $\|a_i\|_\infty^2$ pour diminuer la cohérence (b).

- Nous définissons un $p$-échantillonneur à densité variable comme une trajectoire dont la mesure empirique (ou mesure d'occupation) converge vers $p$ quand le longueur de la trajectoire converge vers l'infini. Nous en donnons deux exemples. Le premier est fondé sur des marches aléatoires (Fig. 1.13(a)). Nous introduisons une inégalité de concentration nouvelle pour une chaîne de Markov à valeurs dans les matrices Hermitiennes (chapitre 3). Celle-ci permet de déduire un résultat d'échantillonnage compressif où le nombre de mesures est $\mathcal{O}\left(\frac{s}{\varepsilon}\cdot\sum_{j=1}^{n}\|a_j\|_\infty^2\cdot\log(6n/\eta)\right)$, où $\varepsilon$ est le trou spectral de la chaîne de Markov. Le second est la solution d'un problème de voyageur de commerce (Fig. 1.13(b)) et nous relions la distribution selon laquelle les « villes » sont tirées et la distribution de la courbe reliant ces points. Le message mis en avant par ces deux approches est qu'en plus de la densité d'échantillonnage, le temps de mélange des processus d'échantillonnage est crucial.

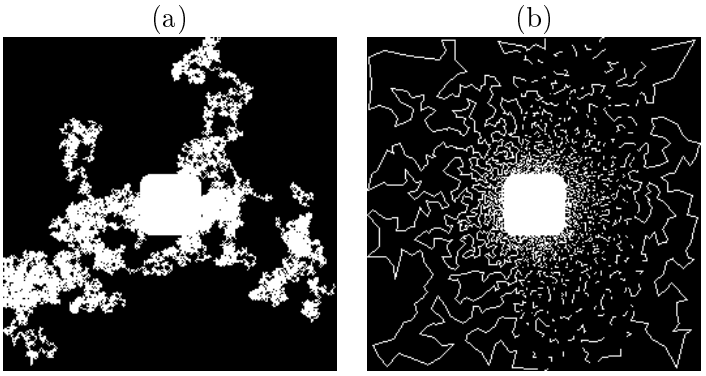


FIGURE 1.13: Exemple d'échantillonneurs à densité variable reposant sur des marches aléatoires (a) et sur la solution d'un problème de voyageur de commerce (b).

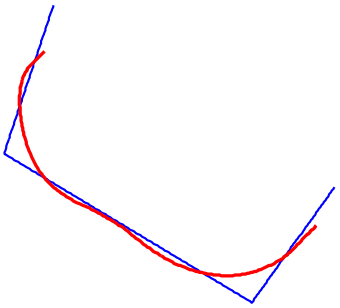

FIGURE 1.14: Exemple de projection d'une trajectoire non physiquement plausible (la courbe bleue parcourue à vitesse constante). En rouge, la solution du problème de projection sur $\mathcal{S}_T$.

- Les trajectoires continues n'étant pas physiquement jouables par un scanner, nous développons un algorithme de projection sur l'ensemble (convexe) des contraintes (chapitre 4). Une illustration sur un exemple synthétique est donné dans la Fig. 1.14.

- En s'appuyant sur l'algorithme précédent, nous proposons un nouvel algorithme de projection sur des espaces de mesures. Le cas intéressant pour l'IRM est l'espace des mesures portées par des courbes admissibles. Nous donnons des illustrations pour des espaces de mesures plus généraux. En particulier, si la mesure que l'on projette est proportionnelle aux niveaux de gris d'une image, il est possible de la représenter par une mesure portée par une courbe, et donc de représenter une image par une courbe (Fig. 1.15) !

- Cet algorithme de projection permet dans le cas de l'IRM de réaliser des schémas d'échantillonnage à densité variable par des courbes admissibles (Fig. 1.16). Les résultats de reconstruction correspondant permettent de dépasser les schémas classiques (spirale, radial) de 3 dB au minimum en termes de SNR.

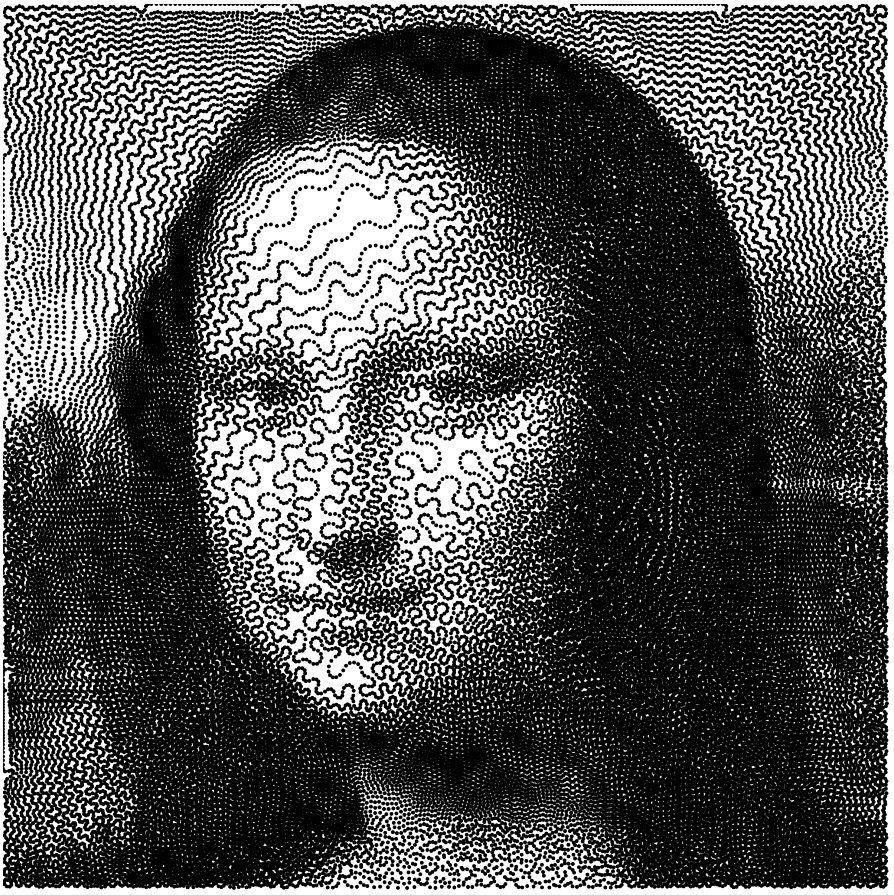

FIGURE 1.15: Représentation de Mona Lisa par un élément de $\mathcal{S}_T$.

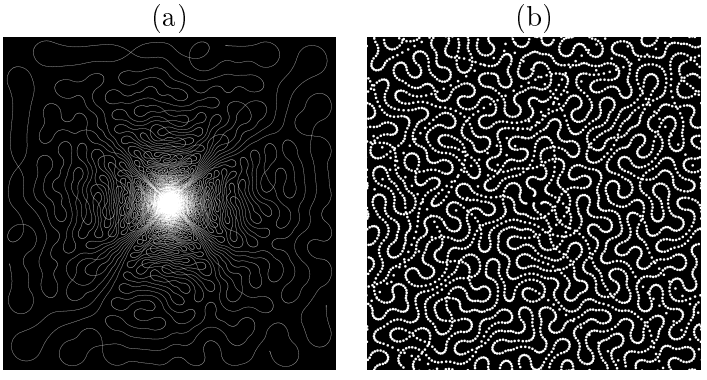


FIGURE 1.16: Projection d'une densité cible à décroissance radiale sur l'ensemble des mesures portées par des courbes de $\mathcal{S}_T$ (a). Zoom sur les basses fréquences (b).

## 1.5 Organisation

L'organisation de cette thèse sur articles suit les 4 travaux suivants:

- Variable Density Sampling with Continuous Trajectories. N Chauffert, P Ciuciu, J Kahn, P Weiss. SIAM Journal on Imaging Sciences 7 (4), 1962-1992.
- A projection algorithm for gradient waveforms design in Magnetic Resonance Imaging. N. Chauffert, P. Weiss, J. Kahn and P. Ciuciu, en révision à IEEE Trans. on Medical Imaging en juin 2015.
- A projection method on measures sets. N. Chauffert, P. Ciuciu, J. Kahn and P. Weiss, soumis en 2015 à Constructive Approximation.
- On the generation of sampling schemes for Magnetic Resonance Imaging. C. Boyer, N. Chauffert, J. Kahn, P. Ciuciu, Pierre Weiss (article soumis prochainement).

Dans le chapitre 2, nous définissons la notion d'échantillonneur à densité variable. Dans le cadre de mesures isolées, cela correspond au théorème 1.3 évoqué précédemment. Nous proposons des échantillonneurs à densité variable *continus*, et montrons que les caractéristiques importantes des schémas d'échantillonnage sont la densité cible d'une part, et la vitesse de recouvrement de l'espace-$k$ d'autre part. Dans le cas des schémas continus reposant sur des marches aléatoires, la quantité qui gouverne la vitesse de recouvrement est le trou spectral: plus il est grand, plus l'espace est parcouru rapidement. Nous introduisons également des schémas d'échantillonnage reposant sur des solutions du voyageurs de commerce qui fournissent de meilleurs résultats en reconstruction, car le processus recouvre plus rapidement l'espace que les marches aléatoires.

Dans le chapitre 3, nous prouvons une inégalité de concentration pour une chaîne de Markov à valeurs dans les matrices Hermitiennes. Cela permet de donner un résultat d'échantillonnage compressif pour un échantillonnage le long d'une marche aléatoire.

Les schémas continus proposés dans le chapitre 2 ne sont toutefois pas implémentables sur des scanners IRM. Par exemple, pour les schémas solution du voyageur de commerce, le parcours de la trajectoire à vitesse constante comporte des singularités au voisinage des « villes ». Le seul parcours implémentable impose que les gradients soient nuls à chacun de ces points critique. Cela induit un temps d'acquisition extrêment long d'une part, et la densité est grandement modifiée d'autre part. Dans le chapitre 4, nous proposons un algorithme de projection d'une trajectoire qui peut être irréaliste (e.g., un parcours à vitesse constante de la solution du problème de voyageur de commerce) sur l'ensemble des contraintes. Un algorithme de descente de gradient proximal accéléré est utilisé pour

résoudre ce problème. Les caractéristiques des trajectoires obtenues présentent de bonnes propriétés pour l'échantillonnage: le temps d'acquisition est fixé et l'écart à la densité d'échantillonnage de la courbe initiale est faible.

Dans le chapitre 5, nous proposons un algorithme de projection sur un espace de mesures portés par des paramétrisations quelconques. Nous montrons que ce problème généralise le problème de *stippling* (représentation d'une image en niveau de gris par des disques noirs sur fond blanc). Ceci correspond au cas de la projection sur des mesures portées par des mesures de Dirac. En ajoutant des contraintes sur les paramétrisations, nous pouvons considérer des projections sur des mesures portées par des trajectoires avec différents types de contraintes (sur la vitesse ou l'accélération, en norme $\ell_1$, $\ell_2$ ou $\ell_\infty$...). En particulier, il est possible de projeter sur des mesures portées par des courbes de $\mathcal{S}_T$.

Dans le chapitre 6, nous appliquons cet algorithme pour projeter des mesures d'échantillonnage 2D utilisées en IRM (à décroissance radiale des basses vers les hautes fréquences). La projection sur des mesures isolées (portées par des mesures de Dirac) donne des meilleurs résultats de reconstruction que l'échantillonnage reposant sur le tirage i.i.d. des échantillons. De même, nous montrons que la projections sur des mesures portées par $\mathcal{S}_T$ donne des meilleurs résultats de reconstruction que les schémas classique (radial et spiral). Enfin, la projection sur des espaces portés par des segments donne des meilleurs résultats que l'échantillonnage radial.

## 1.6 Organization and contributions

Compressed Sensing theory cannot be used directly in MRI for two reasons. First, the incoherence property between the acquisition (Fourier) and sparsifying (Wavelet) basis is not fulfilled. This is the "coherence barrier" (Adcock et al., 2013). Second, most of Compressed Sensing schemes are obtained by drawing randomly and independently the samples (Rauhut, 2010; Candès and Plan, 2011), leading to a sampling scheme of isolated points. Magnetic field gradients impose kinematics constraints on the $k$-space sampling trajectory (the set $\mathcal{S}_T$ defined in Eq. (1.7)). To date, the application of CS to MRI consists of drawing randomly sample locations in a 2D plane according to CS theory, and then acquiring the MR signal along the third (readout) direction (Fig. 1.11). This strategy is sub-optimal from a CS point of view because the sampling density is constant in the readout direction, whereas according to CS theory, the sampling density should decrease from low to high frequencies.

In Chapter 2, we show that a deterministic sampling of the $k$-space center combined with an i.i.d. drawing can break the "coherence barrier" and that an optimal sampling

density can be introduced to theoretically reduce the number of samples to be acquired while allowing perfect recovery (Fig. 1.12(a)). We show that sampling strategies based on empirical sampling distributions can provide better reconstruction results in CS-MRI, since the signal structure is not taken into account in the model: the signal is supposed to be sparse, although no prior knowledge is considered on the locations of non-zero coefficients. We define a *variable density sampler* (VDS) as a process which empirical distribution converges towards a target distribution. This definition encompasses both the previous i.i.d. drawing and continuous sampling procedures. Continuity is a necessary condition for a sampling trajectory to belong to $\mathcal{S}_T$. We give two examples of continuous VDS: random walks and Travelling-Salesman Problem-based (Fig. 1.13(a) and (b)). From these two examples, we illustrate two crucial properties of a VDS: its empirical distribution and its mixing time, i.e., the time to cover the $k$-space.

Chapter 3 is dedicated to introducing a new Hermitian matrix-valued concentration inequality. This inequality is the cornerstone to prove that if the samples are acquired along a Markov chain, the required number of measurements to ensure exacte reconstruction is $\mathcal{O}\left(\frac{s}{\varepsilon}\cdot\sum_{j=1}^{n}\|a_j\|_\infty^2\cdot\log(6n/\eta)\right)$, where $\varepsilon$ is the spectral gap of the chain.

To sum up, in Chapter 2, we described continuous sampling trajectories, which do not belong to $\mathcal{S}_T$ in general. Next, in Chapter 4, we introduce a projection algorithm that projects an initial parameterization onto the set $\mathcal{S}_T$. This method provides admissible gradient waveforms to traverse a curve that is *close to* the initial parameterization. Compared to existing reparameterization methods (Hargreaves et al., 2004; Lustig et al., 2008), the main difference of our algorithm is that the curve support is no longer constrained. The consequence is that the sampling density is better preserved compared to classical reparameterization techniques, and the sampling time is user-defined and thus generally shorter than the one provided by existing methods. An illustration of the projection algorithm is given in Fig. 1.14: the input curve is shown in blue and its projection onto the set $\mathcal{S}_T$ is depicted in red.

In Chapter 5, we develop an algorithm to design a VDS that fulfills the kinematics constraints. The proposed algorithm is actually more generic than the one we push forward in the previous chapter: it allows to project a density $p$ (e.g., the target distribution $\pi$ defined in Fig. 1.12(a)) onto a set of measures $\mathcal{M}_N$. We give necessary and sufficient conditions on a sequence of sets $(\mathcal{M}_N)_{N\in\mathbb{N}}$ such that the solution of the projection problem converges to $p$. In addition, we provide an algorithm to find an approximation of the solution of the measure-projection algorithm for measures carried by parameterizations. Such measures encompasses sums of Dirac and measures carried by (discrete) curves with

kinematics constraints (e.g. $\in \mathcal{S}_T$). This algorithm is generic enough to be deployed beyond the MRI field. For instance, if $p$ is a grayscale image, the solution to the projection problem gives a representation with isolated dots (sum of Dirac) called *stippling*, or with a continuous trajectory (continuous line drawing). To illustrate this point, Mona Lisa is represented by an element of $\mathcal{S}_T$ in Fig. 1.15.

In Chapter 6, we apply this algorithm to design admissible sampling curves in MRI. On retrospective CS simulations, we show that our approach improves reconstruction results compared to spiral or radial sampling. The sampling schemes obtained with this method fulfill the two requirements described in Chapter 2, namely the sampling density with a variable profile and the fast mixing time to cover the $k$-space sufficiently well. A sampling pattern obtained with this method is provided Fig. 1.16(a) and (b): zoom on the low frequencies. We conducted numerical experiments on image sizes of $256 \times 256$ and $2048 \times 2048$.

# Chapter 2

# Variable density sampling with continuous trajectories

This chapter is based on (Chauffert et al., 2014a).

## Abstract

Reducing acquisition time is a crucial challenge for many imaging techniques. Compressed Sensing (CS) theory offers an appealing framework to address this issue since it provides theoretical guarantees on the reconstruction of sparse signals by projection on a low dimensional linear subspace. In this paper, we focus on a setting where the imaging device allows to sense a fixed set of measurements. We first discuss the choice of an optimal sampling subspace allowing perfect reconstruction of sparse signals. Its design relies on the random drawing of independent measurements. We discuss how to select the drawing distribution and show that a mixed strategy involving partial deterministic sampling and independent drawings can help breaking the so-called "coherence barrier". Unfortunately, independent random sampling is irrelevant for many acquisition devices owing to acquisition constraints. To overcome this limitation, the notion of Variable Density Samplers (VDS) is introduced and defined as a stochastic process with a prescribed limit empirical measure. It encompasses samplers based on independent measurements or continuous curves. The latter are crucial to extend CS results to actual applications. We propose two original approaches to design continuous VDS, one based on random walks over the acquisition space, and one based on Traveling Salesman Problem. Following theoretical considerations and retrospective CS simulations in magnetic resonance

imaging, we intend to highlight the key properties of a VDS to ensure accurate sparse reconstructions, namely its limit empirical measure and its mixing time[1].

## 2.1 Introduction

Variable density sampling is a technique that is extensively used in various sensing devices such as magnetic resonance imaging (MRI), in order to shorten scanning time. It consists in measuring only a small number of random projections of a signal/image on elements of a basis drawn according to a given density. For instance, in MRI where measurements consist of Fourier (or more generally $k$-space) coefficients, it is common to sample the Fourier plane center more densely than the high frequencies. The image is then reconstructed from this incomplete information by dedicated signal processing methods. To the best of our knowledge, *variable density sampling* has been proposed first in the MRI context by (Spielman et al., 1995) where spiral trajectories were pushed forward. Hereafter, it has been used in this application (see e.g. (Tsai and Nishimura, 2000; Kim et al., 2003; Park et al., 2005) to quote a few), but also in other applications such as holography (Rivenson et al., 2010; Marim et al., 2010). This technique can hardly be avoided in specific imaging techniques such as radio interferometry or tomographic modalities (e.g., X-ray) where sensing is made along fixed sets of measurements (Wiaux et al., 2009; Sidky et al., 2006).

In the early days of its development, variable density sampling was merely an efficient heuristic to shorten acquisition time. It has recently found a partial justification in the Compressed Sensing (CS) literature. Even though this theory is not yet mature enough to fully explain the practical success of variable density sampling, CS provides good hints on how to choose the measurements (i.e., the density), how the signal/image should be reconstructed and why it works. Let us now recall a typical result emanating from the CS literature for orthogonal systems. A vector $x \in \mathbb{C}^n$ is said $s$-sparse if it contains at most $s$ non-zero entries. Denote by $a_i,\ i \in \{1,\dots,n\}$ the sensing vectors and by $y_i = \langle a_i, x\rangle$ the possible measurements. Typical CS results state that if the signal (or image) $x$ is $s$-sparse and if $\mathbf{A} = \begin{pmatrix} a_1^* \\ \vdots \\ a_n^* \end{pmatrix}$ satisfies an incoherence property (defined in the sequel), then $m = O(s\log(n)^\alpha)$ measurements chosen randomly among the elements of $y = \mathbf{A}x$ are enough to ensure perfect reconstruction of $x$. The constant $\alpha > 0$ depends on additional properties on $x$ and $\mathbf{A}$. The set of actual measurements is denoted $\Omega \subseteq \{1,\dots,n\}$ and $\mathbf{A}_\Omega$ is the matrix formed by selecting a subset of rows of $\mathbf{A}$ in $\Omega$. The reconstruction of $x$

[1] Part of this work is based on the conference proceedings: (Chauffert et al., 2013b; Chauffert et al., 2013a; Chauffert et al., 2013c).

knowing $y_\Omega = \mathbf{A}_\Omega x$ is guaranteed if it results from solving the following $\ell_1$ minimization problem:

$$\min_{z \in \mathbb{C}^n} \|z\|_1 \qquad \text{subject to} \qquad \mathbf{A}_\Omega z = y_\Omega. \tag{2.1}$$

Until recent works (Rauhut, 2010; Juditsky et al., 2011; Candès and Plan, 2011), no general theory for selecting the rows was available. In the latter, the authors have proposed to construct $\mathbf{A}_\Omega$ by drawing $m$ rows of $\mathbf{A}$ at random according to a discrete probability distribution or density $p = (p_1, \dots, p_n)$. The choice of an optimal distribution $p$ is an active field of research (see e.g. (Chauffert et al., 2013b; Krahmer and Ward, 2014; Adcock et al., 2013)) that remains open in many regards.

Drawing independent rows of $\mathbf{A}$ is interesting from a theoretical perspective, however it has little practical relevance since standard acquisition devices come with acquisition constraints. For instance, in MRI, the coefficients are acquired along *piece-wise continuous* curves on the $k$-space. The first paper performing variable density sampling in MRI (Spielman et al., 1995) has fulfilled this constraint by considering spiral sampling trajectories. The standard reference about CS-MRI (Lustig et al., 2007) has proposed to sample the MRI signal along parallel lines in the 3D $k$-space. Though spirals and lines can be implemented easily on a scanner, it is likely that more general trajectories could provide better reconstruction results, or save more scanning time.

The main objective of this paper is to propose new strategies to sample a signal along more general continuous curves. Although continuity is often not sufficient for practical implementation on actual scanner, we believe that it is a first important step towards more physically plausible compressed sampling paradigms. As far as we know, this research avenue is relatively new. The problem was first discussed in (Willett., 2011) and some heuristics were proposed. The recent contributions (Polak et al., 2012b; Bigot et al., 2013) have provided theoretical guarantees when sampling is performed along fixed sets of measurements (e.g. straight lines in the Fourier plane), but have not addressed generic continuous sampling curves yet.

The contributions of this paper are threefold. First, we bring a well mathematically grounded definition of *variable density samplers* and provide various examples. Second, we discuss how the sampling density should be chosen in practice. This discussion mostly relies on variations around the theorems provided in (Rauhut, 2010; Candès and Plan, 2011). In particular, we justify the deterministic sampling of a set of highly coherent vectors to overcome the so-called "coherence barrier". In the MRI case, this amounts to deterministically sampling the $k$-space center. Our third and maybe most impacting contribution is to provide practical examples of variable density samplers along

continuous curves and to derive some of their theoretical properties. These samplers are defined as parametrized random curves that asymptotically fit a target distribution (e.g. the one shown in Fig. 2.1 (a)). More specifically, we first propose a *local* sampler based on random walks over the acquisition space (see Fig. 2.1 (b)). Second, we introduce a *global* sampler based on the solution of a Travelling Salesman Problem amongst randomly drawn "cities" (see Fig. 2.1 (c)). In both cases, we investigate the resulting density. To finish with, we illustrate the proposed sampling schemes on 2D and 3D MRI simulations. The reconstruction results provided by the proposed techniques show that the PSNR can be substantially improved compared to existing strategies proposed e.g. in (Lustig et al., 2007). Our theoretical results and numerical experiments on retrospective CS show that two key features of variable density samplers are the **limit of their empirical measure** and their **mixing properties**.

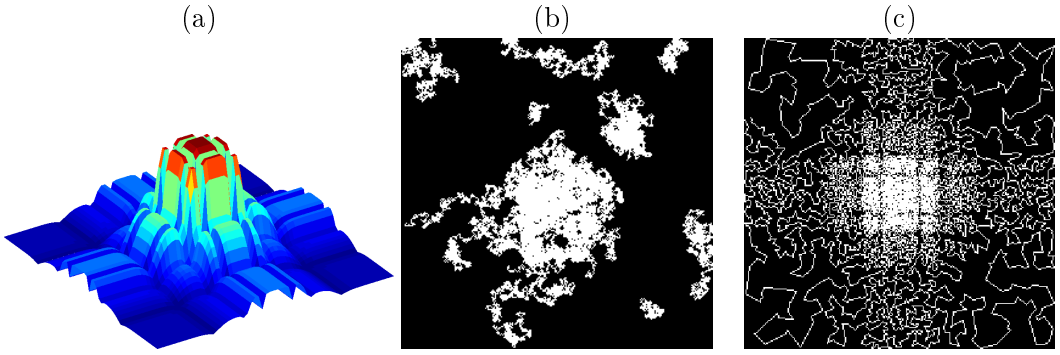


FIGURE 2.1: (a): Target distribution $\pi$. Continuous random trajectories reaching distribution $\pi$ based on Markov chains (b) and on a TSP solution (c).

The rest of this paper is organized as follows. First, we introduce a precise definition of a variable density sampler (VDS) and recall CS results in the special case of independent drawings. Then, we give a closed form expression for the optimal distribution depending on the sensing matrix $\mathbf{A}$, and justify that a partial deterministic sampling may provide better reconstruction guarantees. Hereafter, in Sections 2.3 and 2.4, we introduce two strategies to design continuous trajectories over the acquisition space. We show that the corresponding sampling distributions converge to a target distribution when the curve length tends to infinity. Finally, we demonstrate on simulation results that our TSP-based approach is promising in the MRI context (Section 2.5) since it outperforms its competing alternatives either in terms of PSNR at fixed sampling rate, or in terms of acceleration factor at fixed PSNR.

## Notation

The main definitions used throughout the paper are defined in Tab. 2.1.

TABLE 2.1: General notation used in the paper.

| | Notation | Definition | Domain |
|---|---|---|---|
| Compressed Sensing | $n$ | Acquisition and signal space dimensions | $\mathbb{N}$ |
| | $m$ | Number of measurements | $\mathbb{N}$ |
| | $R = n/m$ | Sampling ratio | $\mathbb{Q}$ |
| | $\mathbf{A}$ | Full orthogonal acquisition matrix | $\mathbb{C}^{n\times n}$ |
| | $\Omega$ | Set of measurements | $\{1,\dots,n\}^m$ |
| | $\mathbf{A}_\Omega$ | Matrix formed with the rows of $\mathbf{A}$ corresponding to indexes belonging to $\Omega$ | $\mathbb{C}^{m\times n}$ |
| | $x$ | Sparse signal | $\mathbb{C}^n$ |
| | $s$ | Number of non zero coefficients of $x$ | $\mathbb{N}$ |
| | $\Delta_n$ | $\left\{p = \begin{pmatrix} p_1 \\ \vdots \\ p_n \end{pmatrix}, 0 \leqslant p_i \leqslant 1, \sum_{i=1}^n p_i = 1\right\}$ | $\mathbb{R}^n$ |
| | $\|\ \|_1$ | $\ell_1$ norm defined for $z \in \mathbb{C}^n$ by $\|z\|_1 = \sum_{i=1}^n \|z_i\|$ | |
| | $\|\ \|_\infty$ | $\ell_\infty$ norm defined for $z \in \mathbb{C}^n$ by $\|z\|_\infty = \max_{1\leqslant i\leqslant n} \|z_i\|$ | |
| MRI application | $k = \begin{pmatrix} k_x \\ k_y \end{pmatrix}$ or $\begin{pmatrix} k_x \\ k_y \\ k_z \end{pmatrix}$ | Fourier frequencies | $\mathbb{R}^2$ or $\mathbb{R}^3$ |
| | $\mathbf{F}_n^*$ | $d$-dimensional discrete Fourier transform on an $n$ pixels image | $\mathbb{C}^{n\times n}$ |
| | $\mathbf{\Psi}_n$ | $d$-dimensional inverse discrete Wavelet transform on an image of $n$ pixels | $\mathbb{C}^{n\times n}$ |
| | | $\mathbf{F}_n^*$ and $\mathbf{\Psi}_n$ are denoted $\mathbf{F}^*$ and $\mathbf{\Psi}$ if no ambiguity | |
| VDS | $\Xi$ | A measurable space which is typically $\{1,\dots,n\}$ or $[0,1]^d$ | |
| | $\mathcal{H}$ | The unit cube $[0,1]^d$ | |
| | $p$ | A probability measure defined on $\Xi$ | |
| | $p(f)$ | $= \int_{x\in\Xi} f(x)\ \mathrm{d}p(x)$, for $f$ continuous and bounded | $\mathbb{R}$ |
| | $\lambda_{[0,1]}$ | The Lebesgue measure on the interval $[0,1]$ | |
| | $X = (X_n)_{n\in\mathbb{N}^*}$ | A time-homogeneous Markov chain on the state space $\{1,\dots,n\}$ | $\{1,\dots n\}^{\mathbb{N}^*}$ |
| | $\mathbf{P}$ | $:= (P_{ij})_{1\leqslant i,j\leqslant n}$ the transition matrix: $P_{ij} := \mathbb{P}(X_k = j \mid X_{k-1} = i), \forall k > 1$ | $\mathbb{R}^{n\times n}$ |
| | $\lambda_i(\mathbf{P})$ | The ordered eigenvalues of $\mathbf{P}$: $1 = \lambda_1(\mathbf{P}) \geqslant \dots \geqslant \lambda_n(\mathbf{P}) \geqslant -1$ | $[-1,1]$ |
| | $\epsilon(\mathbf{P})$ | $= 1 - \lambda_2(\mathbf{P})$, the spectral gap of $\mathbf{P}$ | $[-1,1]$ |
| | $F$ | A set of points $\subset \mathcal{H}$ | $\mathcal{H}^N$ |
| | $C(F)$ | The shortest Hamiltonian path (TSP) amongst points of set $F$ | $\subset \mathcal{H}$ |
| | $T(F,\mathcal{H})$ | The length of $C(F)$ | $\mathbb{R}_+$ |
| | $T(F,R)$ | For any set $R \subseteq \mathcal{H}$, $T(F,R) := T(F \cap R, \mathcal{H})$ | $\mathbb{R}_+$ |

## 2.2 Variable density sampling and its theoretical foundations

To the best of our knowledge, there is currently no rigorous definition of variable density sampling. Hence, to fill this gap, we provide a precise definition below.

**Definition 2.1.** Let $p$ be a probability measure defined on a measurable space $\Xi$. A stochastic process $X = \{X_i\}_{i\in\mathbb{N}}$ or $X = \{X_t\}_{t\in\mathbb{R}_+}$ on state space $\Xi$ is called a *$p$-variable density sampler* if its empirical measure (or occupation measure) weakly converges to $p$ almost surely, that is:

$$\frac{1}{N}\sum_{i=1}^{N} f(X_i) \to p(f) \qquad a.s.$$

or

$$\frac{1}{T}\int_{t=0}^{T} f(X_t)\mathrm{d}t \to p(f) \qquad a.s.$$

for all continuous bounded $f$.

**Example 2.1.** *In the case where $X = (X_i)_{i\in\mathbb{N}}$ is a discrete time stochastic process with discrete state space $\Xi = \{1,\dots,n\}$, definition 2.1 can be slightly simplified. Let us set $Z_j^N = \frac{1}{N}\sum_{i=1}^{N} \mathbf{1}_{X_i=j}$. The random variable $Z_j^N$ represents the proportion of points that fall on position $j$. Let $p$ denote a discrete probability distribution function. Using these notations, $X$ is a $p$-variable density sampler if:*

$$\lim_{N\to+\infty} Z_j^N = p_j \qquad a.s.$$

*In particular, if $(X_i)_{i\in\mathbb{N}}$ are i.i.d. samples drawn from $p$, then $X$ is a $p$-variable density sampler. This simple example is the most commonly encountered in the compressed sensing literature and we will review its properties in paragraph 2.2.1.*

**Example 2.2.** *More generally, drawing independent random variables according to distribution $p$ is a VDS if the space $\Xi$ is second countable, owing to the strong law of large numbers.*

**Example 2.3.** *An irreducible aperiodic Markov chain on a finite sample space is a VDS for its stationary distribution (or invariant measure); see Section 2.3.3.*

**Example 2.4.** *In the deterministic case, for a dynamical system, definition 2.1 closely corresponds to the ergodic hypothesis, that is time averages are equal to expectations over space. We discuss an example that makes use of the TSP solution in section 2.4.*

The following proposition directly relates the VDS concept to the time spent by the process in a part of the space, as an immediate consequence of the porte-manteau lemma (see e.g. (Billingsley, 2009)).

**Proposition 2.2.** *Let $p$ denote a Borel measure defined on a set $\Xi$. Let $B \subseteq \Xi$ be a measurable set. Let $X : \mathbb{R}_+ \to \Xi$ (resp. $X : \mathbb{N} \to \Xi$) be a stochastic process. Let $\mu$*

*denote the Lebesgue measure on* $\mathbb{R}$. *Define* $\mu_X^t(B) = \frac{1}{t}\mu(\{s \in [0,t],\ X(s) \in B\})$ *(resp.* $\mu_X^n(B) = \frac{1}{n}\sum_{i=1}^n \mathbf{1}_{X(i)\in B}$*). Then, the following two propositions are equivalent:*

$$
\begin{aligned}
&(i) && X \text{ is a } p\text{-VDS} \\
&(ii) && \text{Almost surely, } \forall B \subseteq \Xi \text{ a Borel set with } p(\partial B) = 0, \\
& && \lim_{t\to+\infty} \mu_X^t(B) = p(B) \qquad a.s. \\
&(resp.) && \lim_{n\to+\infty} \mu_X^n(B) = p(B) \qquad a.s.
\end{aligned}
$$

*Remark* 2.3. Definition 2.1 is a generic definition that encompasses both discrete and continuous time and discrete and continuous state space since $\Xi$ can be any measurable space. In particular, the recent CS framework on orthogonal systems (Rauhut, 2010; Candès and Plan, 2011) falls within this definition.

Definition 2.1 does not encompass some useful sampling strategies. We propose a definition of a generalized VDS, which encompasses stochastic processes indexed over a bounded time set.

**Definition 2.4.** A sequence $\{\{X_t^{(n)}\}_{0\leqslant t\leqslant T_n}\}_{n\in\mathbb{N}}$ is a *generalized p-VDS* if the sequence of occupation measures converges to $p$ almost surely, that is:

$$\frac{1}{T_n}\int_{t=0}^{T_n} f(X_t^{(n)})\mathrm{d}t \to p(f) \qquad a.s.$$

*Remark* 2.5. Let $(X_t)_{t\in\mathbb{R}}$ be a VDS, and $(T_n)_{n\in\mathbb{N}}$ be any positive sequence such that $T_n \to \infty$. Then the sequence defined by $X_t^{(n)} = X_t$ for $0 \leqslant t \leqslant T_n$ is a generalized VDS.

**Example 2.5.** *Let* $\Xi = \mathbb{R}^2$, *and consider* $r : [0,1] \mapsto \mathbb{R}^+$ *a strictly increasing smooth function. We denote by* $r^{-1} : [r(0), r(1)] \to \mathbb{R}$ *its inverse function and by* $\dot{r^{-1}}$ *the derivative of* $r^{-1}$. *Consider a sequence of spiral trajectories* $s_N : [0,N] \to \mathbb{R}^2$ *defined by* $s_N(t) = r\left(\frac{t}{N}\right)\begin{pmatrix}\cos(2\pi t)\\ \sin(2\pi t)\end{pmatrix}$. *Then* $s_N$ *is a generalized VDS for the distribution* $p$ *defined by:*

$$p(x,y) = \begin{cases} \frac{\dot{r^{-1}}\left(\sqrt{x^2+y^2}\right)}{2\pi\int_{\rho=r(0)}^{r(1)} \dot{r^{-1}}(\rho)\rho\mathrm{d}\rho} & \text{if} \quad r(0) \leqslant \sqrt{x^2+y^2} \leqslant r(1) \\ 0 & \text{otherwise} \end{cases}$$

*A simple justification is that the time spent by the spiral in the infinitesimal ring* $\{(x,y) \in \mathbb{R}^2, \rho \leqslant \sqrt{x^2+y^2} \leqslant \rho + \mathrm{d}\rho\}$ *is* $\int_{r^{-1}(\rho)}^{r^{-1}(\rho+\mathrm{d}\rho)} \mathrm{d}t \propto \dot{r^{-1}}(\rho)$.

### 2.2.1 Theoretical foundations - Independent VDS

CS theories provide strong theoretical foundations of VDS based on independent drawings. In this paragraph, we recall a typical result that motivates independent drawing in the $\ell_1$ recovery context (Rauhut, 2010; Foucart and Rauhut, 2013; Candès and Plan, 2011; Krahmer and Ward, 2014; Chauffert et al., 2013b; Bigot et al., 2013; Adcock et al., 2013). Using the notation defined in the introduction, let us give a slightly modified version of (Rauhut, 2010, Theorem 4.2).

**Theorem 2.6.** *Let $p = (p_1, \dots, p_n)$ denote a probability distribution on $\{1, \dots, n\}$ and $\Omega \subset \{1, \dots, n\}$ denote a random set obtained by $m$ independent drawings with respect to distribution $p$. Let $S \in \{1, \dots, n\}$ be an arbitrary set of cardinality $s$. Let $x$ be an $s$-sparse vector with support $S$ such that the signs of its non-zero entries is a Rademacher or Steinhaus sequence*[2]*. Define:*

$$K(\mathbf{A}, p) := \max_{k \in \{1 \dots n\}} \frac{\|a_k\|_\infty^2}{p_k} \tag{2.2}$$

*Assume that:*

$$m \geqslant C K(\mathbf{A}, p) s \ln^2\left(\frac{6n}{\eta}\right) \tag{2.3}$$

*where $C \approx 26.25$ is a constant. Then, with probability $1 - \eta$, vector $x$ is the unique solution of the $\ell_1$ minimization problem (2.1).*

*Remark* 2.7. Candès and Plan have stated stronger results in the case of real matrices in (Candès and Plan, 2011). Namely, the number of necessary measurements was decreased to $O(s \log(n))$, with lower constants and without any assumption on the vector signs. Their results have been derived using the so-called "golfing scheme" proposed in (Gross, 2011). It is likely that these results could be extended to the complex case, however it would not change the optimal distribution which is the main point of this paper. We thus decided to stick to Theorem 2.6.

The choice of an accurate distribution $p$ is crucial since it directly impacts the number of measurements required. In the MRI community, a lot of heuristics have been proposed so far to identify the *best* sampling distribution. In the seminal paper on CS-MRI (Lustig et al., 2007), Lustig *et al* have proposed to sample the $k$-space using a density that polynomially decays towards high frequencies. More recently, Knoll *et al* have generalized this approach by inferring the best exponent from MRI image databases (Knoll et al., 2011). It is actually easy to derive the theoretically *optimal distribution*, i.e. the one

[2] A Rademacher (resp. Steinhaus) random variable is uniformly distributed on $\{-1; 1\}$ (resp. on the torus $\{z \in \mathbb{C}; |z| = 1\}$).

that minimizes the right hand-side in (2.3) as shown in Proposition 2.8, introduced in (Chauffert et al., 2013b).

**Proposition 2.8.** *Denote by* $K^*(\mathbf{A}) := \min_{p \in \Delta_n} K(\mathbf{A}, p)$.

*(i) the optimal distribution* $\pi \in \Delta_n$ *that minimizes* $K(\mathbf{A}, p)$ *is:*

$$\pi_i = \frac{\|a_i\|_\infty^2}{\sum_{i=1}^n \|a_i\|_\infty^2} \tag{2.4}$$

*(ii)* $K^*(\mathbf{A}) = K(\mathbf{A}, \pi) = \sum_{i=1}^n \|a_i\|_\infty^2$.

*Proof.* (i) Taking $p = \pi$, we get $K(\mathbf{A}, \pi) = \sum_{i=1}^n \|a_i\|_\infty^2$. Now assume that $q \neq \pi$, since $\sum_{k=1}^n q_k = \sum_{k=1}^n \pi_k = 1$, $\exists j \in \{1, \dots, n\}$ such that $q_j < \pi_j$. Then $K(\mathbf{A}, q) \geqslant \|a_j\|_\infty^2 / q_j > \|a_j\|_\infty^2 / \pi_j = \sum_{i=1}^n \|a_i\|_\infty^2 = K(\mathbf{A}, \pi)$. So, $\pi$ is the distribution that minimizes $K(\mathbf{A}, p)$.

(ii) This equality is a consequence of $\pi$'s definition.

□

The theoretical optimal distribution only depends on the acquisition matrix, i.e. on the acquisition and sparsifying bases. For instance, if we measure some Fourier frequencies of a sparse signal in the time domain (a sum of diracs), we should sample the frequencies according to a uniform distribution, since $\|a_i\|_\infty = 1/\sqrt{n}$ for all $1 \leqslant i \leqslant n$. In this case, $K^*(\mathbf{F}) = 1$ and the number of measurements $m$ is proportional to $s$, which is in accordance with the seminal paper by Candès *et al.* (Candès et al., 2006a).

### Independent drawings in MRI

In the MRI case, the images are usually assumed sparse (or at least compressible) in a wavelet basis, while the acquisition is performed in the Fourier space. In this setting, the acquisition matrix can be written as $\mathbf{A} = \mathbf{F}^*\mathbf{\Psi}$. In that case, the *optimal distribution* only depends on the choice of the wavelet basis. The optimal distributions in 2D and 3D are depicted in Fig. 2.2(a)-(b), respectively if we assume that the MR images are sparse in the Symmlet basis with 3 decomposition levels in the wavelet transform.

Let us mention that similar distributions have been proposed in the literature. First, an alternative to independent drawing was proposed by Puy *et al.* (Puy et al., 2011). Their approach consists in selecting or not a frequency by drawing a Bernoulli random variable.

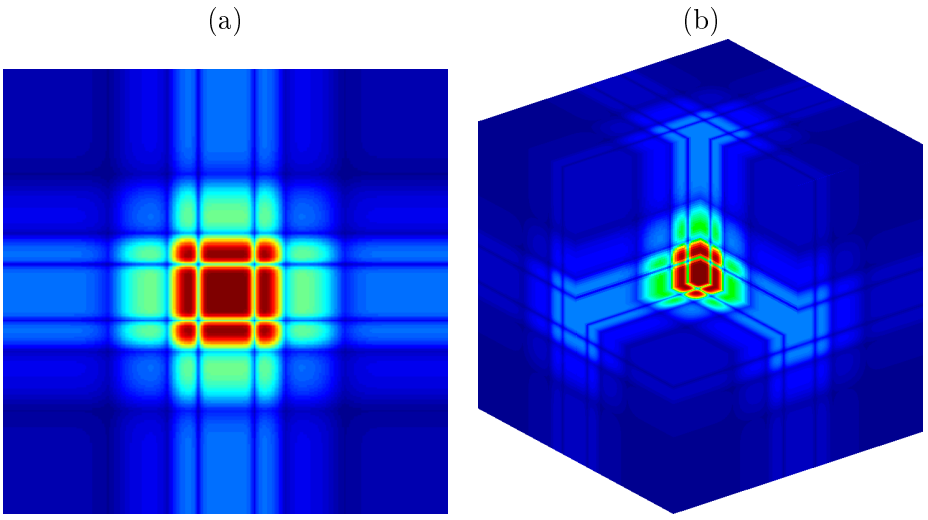


FIGURE 2.2: Optimal distribution $\pi$ for a Symmlet-10 tranform in 2D (a) and a maximal projection of the optimal distribution in 3D (b).

Its parameter is determined by minimizing a quantity that slightly differs from $K(\mathbf{A}, p)$. Second, Krahmer and Ward (Krahmer and Ward, 2014) tried to unify theoretical results and empirical observations in the MRI framework. For Haar wavelets, they have shown that a polynomial distribution on the 2D $k$-space which varies as $1/(k_x^2 + k_y^2)$ is close to the optimal solution since it verifies $K(\mathbf{A}, p) = O(\log(n))$. Our numerical experiments have confirmed that a decay as a power of 2 is near optimal in 2D.

In the next section, we improve the existing theories by showing that a deterministic sampling of highly coherent vectors (i.e. those satisfying $\|a_i\|_\infty^2 \gg \frac{1}{n}$) may decrease the total number of required measurements. In MRI, this amounts to fully sampling the low frequencies, which exactly matches what has been done heuristically hitherto.

### 2.2.2 Mixing deterministic and independent samplings

In a recent work (Chauffert et al., 2013b), we observed and partially justified the fact that a deterministic sampling of the low frequencies in MRI could drastically improve reconstruction quality. The following theorem proven in Appendix 1 provides a theoretical justification to this approach.

**Theorem 2.9.** *Let $S \in \{1, \dots, n\}$ be a set of cardinality $s$. Let $x$ be an $s$-sparse vector with support $S$ such that the signs of its non-zero entries is a Rademacher or Steinhaus sequence. Define the acquisition set $\Omega \subseteq \{1, \dots n\}$ as the union of:*

*(i) a deterministic set $\Omega_1$ of cardinality $m_1$.*

*(ii) a random set* $\Omega_2$ *obtained by* $m_2$ *independent drawing according to distribution* $p$ *defined on* $\{1 \dots n\} \setminus \Omega_1$.

*Denote* $m = m_1 + m_2$, $\Omega_1^c = \{1, \dots, n\} \setminus \Omega_1$ *and let* $\Omega = \Omega_1 \cup \Omega_2$. *Assume that:*

$$m \geqslant m_1 + CK(\mathbf{A}_{\Omega_1^c}, p) s \ln^2\left(\frac{6n}{\eta}\right) \tag{2.5}$$

*where* $C = 7/3$ *is a constant, and* $K(\mathbf{A}_{\Omega_1^c}, p) = \max_{i \in \{1,\dots,n\} \setminus \Omega_1} \frac{\|a_i\|_\infty^2}{p_i}$. *Then, with probability* $1 - \eta$, *vector* $x$ *is the unique solution of the* $\ell_1$ *minimization problem* (2.1).

This result implies that there exists an optimal partition between deterministically and randomly selected samples, which is moreover easy to compute. For example, consider the optimal distribution $p_i \propto \|a_i\|_\infty^2$, then $K^*(\mathbf{A}_{\Omega_1^c}) = \sum_{i \in \{1,\dots,n\} \setminus \Omega_1} \|a_i\|_\infty^2$. If the measurement matrix contains rows with large values of $\|a_i\|_\infty$, we notice from inequality (2.5) that these frequencies should be sampled deterministically, whereas the rest of the measurements should be obtained from independent drawings. This simple idea is another way of overcoming the so-called coherence barrier (Krahmer and Ward, 2014; Adcock et al., 2013).

A striking example raised in (Bigot et al., 2013) is the following. Assume that $\mathbf{A} = \begin{pmatrix} 1 & 0 \\ 0 & \mathbf{F}_{n-1}^* \end{pmatrix}$. The assumed optimal independent sampling strategy would consist in independently drawing the rows with distribution $p_1 = 1/2$ and $p_k = 1/\sqrt{n-1}$ for $k \geqslant 2$. According to Theorem 2.6, the number of required measurements is $2Cs \ln^2\left(\frac{6n}{\eta}\right)$. The alternative approach proposed in Theorem 2.9 basically performs a deterministic drawing of the first row combined with an independent uniform drawing over the remaining rows. In total, this scheme requires $1 + Cs \ln^2\left(\frac{6n}{\eta}\right)$ measurements and thus reduces the number of measurements by almost a factor 2. Note that the same gain would be obtained by using independent drawings with rejection.

### Mixed deterministic and independent sampling in MRI

In our experiments, we will consider wavelet transforms with three decomposition levels and the Symmlet basis with 10 vanishing moments. Fig. 2.3(a)-(b) shows the modulus of $\mathbf{A}$'s entries with a specific reordering in (b) according to decaying values of $\|a_i\|_\infty$. This decay is illustrated in Fig. 2.3(c). We observe that a typical acquisition matrix in MRI shows large differences between its $\|a_i\|_\infty$ values. More Precisely, there is a small number

of rows with a large infinite norm, sticking perfectly to the framework of Theorem 2.9. This observation justifies the use of a partial deterministic $k$-space sampling, which had already been used in (Lustig et al., 2007; Chauffert et al., 2013b). In Fig. 2.3(d), the set $\Omega_1$ is depicted for a fixed number of deterministic samples $m_1$, by selecting the rows with the largest infinite norms.

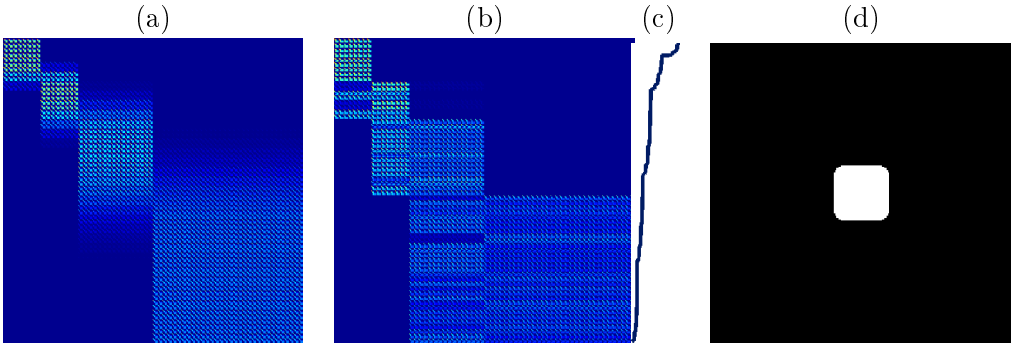


FIGURE 2.3: (a): Absolute magnitudes of $\mathbf{A}$ for a 2D Symmlet basis with 10 vanishing moments and 3 levels of decomposition. (b): same quantities as in (a) but sorted by decaying $\|a_i\|_\infty$ (i.e. by decreasing order). (c): decay of $\|a_i\|_\infty$. (d): Set $\Omega_1$ depicted in the 2D $k$-space.

Hereafter, the strategy we adopt is driven by the previous remarks. All our sampling schemes are performed according to Theorem. 2.9: a deterministic part is sampled, and a VDS is performed on the rest of the acquisition space (*e.g.* the high frequencies in MRI).

## 2.3 Variable density samplers along continuous curves

### 2.3.1 Why independent drawing can be irrelevant

In many imaging applications, the number of samples is of secondary importance compared to the time spent to collect the samples. A typical example is MRI, where the important variable to control is the scanning time. It depends on the total length of the pathway used to visit the $k$-space rather than the number of collected samples. MRI is not an exception and many other acquisition devices have to meet such physical constraints amongst which are scanning probe microscopes, ultrasound imaging, ecosystem monitoring, radio-interferometry or sampling using vehicles subject to kinematic constraints (Willett., 2011). In these conditions, measuring isolated points is not relevant and existing practical CS approaches consist in designing parametrized curves performing a variable density sampling. In what follows, we first review existing variable density

sampling approaches based on continuous curves. Then, we propose two original contributions and analyze some of their theoretical properties. We mostly concentrate on continuity of the trajectory which is not sufficient for implementability in many applications. For instance, in MRI the actual requirement for a trajectory to be implementable is piecewise smoothness. More realistic constraints are discussed in Section 2.6.

### 2.3.2 A short review of samplers along continuous trajectories

The prototypical variable density samplers in MRI were based on spiral trajectories (Spielman et al., 1995). Similar works investigating different shapes and densities from a heuristic point of view were proposed in (Tsai and Nishimura, 2000; Kim et al., 2003; Park et al., 2005). The first reference to compressed sensing appeared in the seminal paper (Lustig et al., 2007). In this work, Lustig *et al* have proposed to perform independent drawings in a 2D plane (defined by the partition and phase encoding directions) and sample continuously along the orthogonal direction to design piecewise continuous schemes in the 3D $k$-space (see Fig. 2.4). These authors have also suggested to make use of randomly perturbed spirals. The main advantage of these schemes lies in their simplicity of practical implementation since they only require minor modifications of classical MRI acquisition sequences.

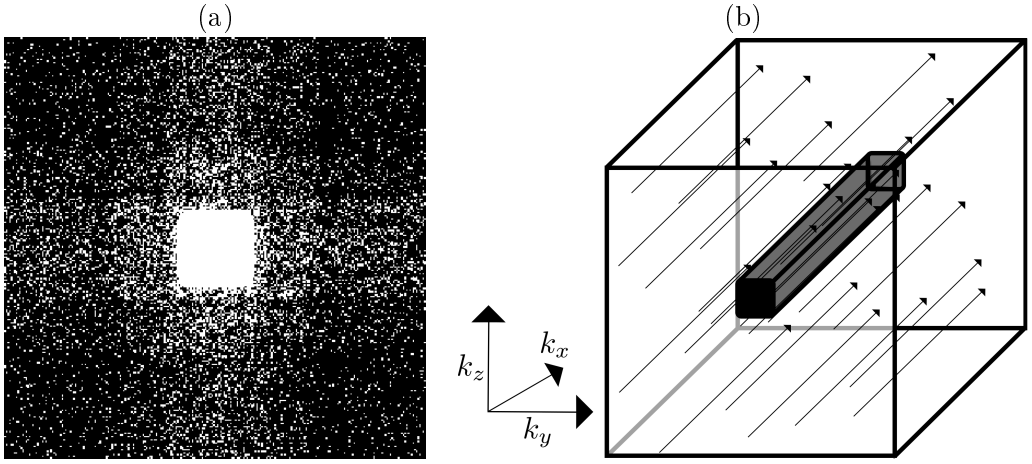


FIGURE 2.4: Classical CS-MRI strategy. (a): 2D independent sampling according to a distribution $\pi$. (b): measurements performed in the orthogonal readout direction.

Recent papers (Polak et al., 2012a; Bigot et al., 2013; Boyer et al., 2014) have generalized CS results from independent drawing of isolated measurements to independent drawings of blocks of measurements. In these contributions, the blocks can be chosen arbitrarily and may thus represent continuous trajectories. Interestingly, these authors have provided closed form expressions for the optimal distribution on the block set. Nevertheless, this distribution is very challenging to compute in large scale problems. Moreover, the

restriction to sets of admissible blocks reduces the versatility of many devices such as MRI and can therefore impact the image reconstruction quality.

In many applications the length of the sampling trajectory is more critical than the number of acquired samples, therefore, finding the shortest pathway amongst random points drawn independently has been studied as a way of designing continuous trajectories (Willett., 2011; Wang et al., 2012). Since this problem is NP-hard, one usually resorts to a TSP solver to get a reasonable suboptimal trajectory. To the best of our knowledge, the only practical results obtained using the TSP were given by Wang *et al* (Wang et al., 2012). In this work, the authors did not investigate the relationship between the initial sample locations and the empirical measure of the TSP curve. In Section 2.4, it is shown that this relationship is crucial to make efficient TSP-based sampling schemes.

In what follows, we first introduce an original sampler based on random walks on the acquisition space and then analyse its asymptotic properties. Our theoretical investigations together with practical experiments allows us to show that the VDS mixing properties play a central role to control its efficiency. This then motivates the need for more global VDS schemes.

### 2.3.3 Random walks on the acquisition space

Perhaps the simplest way to transform independent random drawings into continuous random curves consists in performing random walks on the acquisition space. Here, we discuss this approach and provide a brief analysis of its practical performance in the discrete setting. Through both experimental and theoretical results, we show that this technique is doomed to fail. However, we believe that this theoretical analysis provides a deep insight on what VDS properties characterize its performance.

Let us consider a time-homogeneous Markov chain $X = (X_n)_{n\in\mathbb{N}}$ on the set $\{1,\dots,n\}$ and its transition matrix denoted $\mathbf{P} \in \mathbb{R}^{n\times n}$. If $X$ possesses a stationary distribution, i.e. a row vector $p \in \mathbb{R}^n$ such that $p = p\mathbf{P}$ then, by definition, $X$ is a $p$-variable density sampler.

#### 2.3.3.1 Construction of the transition matrix P

A classical way to design a transition kernel ensuring that (i) $p$ is the stationary distribution of the chain and (ii) the trajectory defined by the chain is continuous, is the Metropolis algorithm (Hastings, 1970). For a pixel/voxel position $i$ in the 2D/3D acquisition space, let us define by $\mathcal{N}(i) \subseteq \{1,\dots,n\}$ its neighbourhood, *i.e.* the set of possible measurement locations allowed when staying on position $i$. Let $|\mathcal{N}(i)|$ denote

the cardinal of $\mathcal{N}(i)$ and define the *proposal kernel* $\mathbf{P}^*$ as $\mathbf{P}^*_{i,j} = |\mathcal{N}(i)|^{-1}\delta_{j\in\mathcal{N}(i)}$. The Metropolis algorithm proceeds as follows:

1. from state $i$, draw a state $i^*$ with respect to the distribution $\mathbf{P}^*_{i,:}$.
2. accept the new state $i^*$ with probability:

$$q(i, i^*) = \min\left(1, \frac{p(i^*)\mathbf{P}^*_{i^*,i}}{p(i)\mathbf{P}^*_{i,i^*}}\right). \tag{2.6}$$

   Otherwise stay in state $i$.

The transition matrix $\mathbf{P}$ can then be defined by $\mathbf{P}_{i,j} = q(i,j)\mathbf{P}^*_{i,j}$ for $i \neq j$. The diagonal is defined in a such a way that $\mathbf{P}$ is a stochastic matrix. It is easy to check that $p$ is an invariant distribution for this chain[3]. It is worth noticing that if the chain is irreducible positive recurrent (which is fulfilled if the graph is connected and the distribution $p$ positive), the ergodic theorem ensures that $X$ is a $p$-VDS.

Unfortunately, trajectories designed by this technique leave huge parts of the acquisition space unexplored (see Fig. 2.5 (a)). To circumvent this problem, we may allow the chain to *jump* to independent locations over the acquisition space. Let $\tilde{\mathbf{P}}$ be the Markov kernel corresponding to independent drawing with respect to $p$, *i.e.* $\tilde{\mathbf{P}}_{i,j} = p_j$ for all $1 \leqslant i, j \leqslant n$. Define:

$$\mathbf{P}^{(\alpha)} = (1-\alpha)\mathbf{P} + \alpha\tilde{\mathbf{P}} \qquad \forall\, 0 \leqslant \alpha \leqslant 1. \tag{2.7}$$

Then the Markov chain associated with $\mathbf{P}^{(0)}$ corresponds to a continuous random walk, while the Markov chain associated with $\mathbf{P}^{(\alpha)}$, $\alpha > 0$ has a nonzero *jump* probability. This means that the trajectory is composed of continuous parts of average length $1/\alpha$.

#### 2.3.3.2 Example

In Fig. 2.5, we show illustrations in the 2D MRI context where the discrete $k$-space is of size $64 \times 64$. On this domain, we set a distribution $p$ which matches distribution $\pi$ in Fig. 2.2 (a). We perform a random walk on the acquisition space until 10% of the coefficients are selected. In Fig. 2.5(a), we set $\alpha = 0$ whereas $\alpha = 0.1$ in Fig. 2.5(b). As expected, $\alpha = 0$ leads to a sampling pattern where large parts of the $k$-space are left unvisited. The phenomenon is partially corrected using a nonzero value of $\alpha$.

[3] If the neighboring system is such that the corresponding graph is connected, then the invariant distribution is unique.

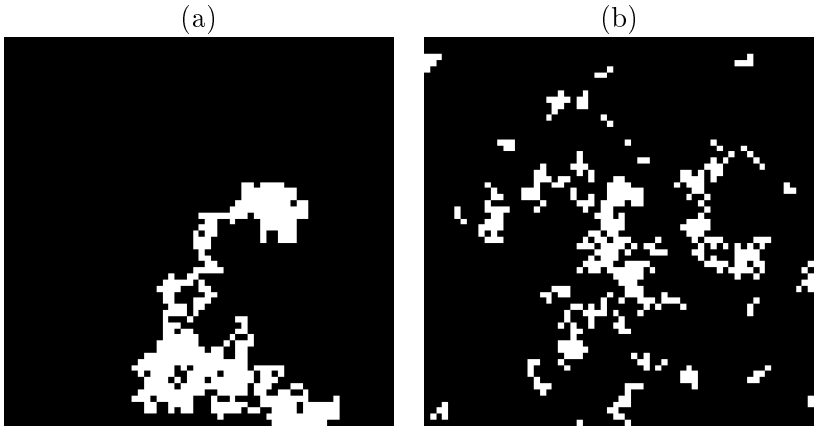


FIGURE 2.5: Example of sampling trajectories in 2D MRI. (a) (resp. (b)): 2D sampling scheme of the $k$-space with $\alpha = 0$ (resp. $\alpha = 0.1$). Drawings are performed until 10% of the coefficients are selected ($m = 0.1n$).

*Remark* 2.10. Performing $N$ iterations of the Metropolis algorithm requires $O(N)$ computations leading to a fast sampling scheme design procedure. In our experiments, we iterate the algorithm until $m$ *different* measurements are probed. Therefore, the number of iterations $N$ required increases non linearly with respect to $m$, and can be time consuming especially when $R = m/n$ is close to 1. This is not a tough limitation of the method since the sampling scheme is computed off-line.

#### 2.3.3.3 Compressed sensing results

Let us assume[4] that $\mathbb{P}(X_1 = i) = p_i$ and that $X_i$ is drawn using $\mathbf{P}$ as a transition matrix. The following result provides theoretical guarantees about the performance of the VDS $X$:

**Proposition 2.11** (see (Chauffert et al., 2013c)). *Let $\Omega := X_1, \dots, X_m \subset \{1, \dots, n\}$ denote a set of $m$ indexes selected using the Markov chain $X$.*

*Then, with probability $1 - \eta$, if*

$$m \geqslant \frac{12}{\epsilon(\mathbf{P})} K^2(\mathbf{A}, p) s^2 \log(2n^2/\eta), \tag{2.8}$$

*every $s$-sparse signal $x$ is the unique solution of the $\ell_1$ minimization problem.*

The proof of this proposition is given in Appendix 2. Before going further, some remarks may be useful to explain this theoretical result.

[4] By making this assumption, there is no burn-in period and the chain $X$ converges more rapidly to its stationary distribution $p$.

*Remark* 2.12. Since the constant $K^2(\mathbf{A}, p)$ appears in Eq. (2.8), the optimal sampling distribution using Markov chains is also distribution $\pi$, as proven in Proposition 2.8.

*Remark* 2.13. In contrast to Theorem 2.6, Proposition 2.11 provides uniform results, i.e. results that hold for *all* $s$-sparse vectors.

*Remark* 2.14. Ineq. 2.8 suffers from the so-called *quadratick bottleneck* (i.e. an $O(s^2 \log(n))$ bound). It is likely that this bound can be improved to $O(s \log(n))$ by developing new concentration inequalities on matrix-valued Markov chains.[5]

*Remark* 2.15. More importantly, it seems however unlikely to avoid the spectral gap $O(1/\epsilon(P))$ using the standard mechanisms for proving compressed sensing results. Indeed, all concentration inequalities obtained so far on Markov chains (see e.g. (Lezaud, 1998; Kargin, 2007; Paulin, 2012a)) depend on $1/\epsilon(P)$. The spectral gap satisfies $0 < \epsilon(P) \leqslant 1$ and corresponds to mixing properties of the chain. The closer the spectral gap to 1, the fastest ergodicity is achieved. Roughly speaking, if $|i - j| > 1/\epsilon(P)$ then $X_i$ and $X_j$ are almost independent random variables. Unfortunately, the spectral gap usually depends on the dimension $n$ (Diaconis and Stroock, 1991). In our example, it can be shown using Cheeger's inequality that $\epsilon(P) = O\left(n^{-\frac{1}{d}}\right)$ if the stationary distribution $\pi$ is uniform (see Appendix 3). This basically means that the number of measurements necessary to accurately reconstruct $x$ could be as large as $O(sn^{1/d} \log(n))$, which strongly limits the interest of this CS approach. The only way to lower this number consists in frequently jumping since Weyl's theorem (Horn and Johnson, 1991) ensures that $\epsilon(P^{(\alpha)}) > \alpha$.

To sum up, the main drawback of random walks lies in their inability to cover the acquisition space quickly since they are based on local considerations. Keeping this in mind, it makes sense to focus on more global displacement strategies that allow a faster exploration of the whole acquisition domain. In the next section, we thus introduce this global sampling alternative based on TSP-solver. Our main contribution is the derivation of the link between a prescribed a priori sampling density and the distribution of samples located on the TSP solution so as to eventually get a VDS.

## 2.4 Travelling salesman-based VDS

In order to design continuous trajectories, we may think of picking points at random and join them using a travelling salesman problem (TSP) solver. Hereafter, we show how to draw the initial points in order to reach a target distribution $p$. In this section, the probability distribution $p$ is assumed to be a density.

[5] In chapter 3, we provide an improvement of the theorem that breaks the quadratick bottleneck (see proposition 3.11, page 76).

### 2.4.1 Introduction

The naive idea would consist in drawing some points according to the distribution $p$ and joining them using a TSP solver. Unfortunately, the trajectory which results from joining all samples does not fit the distribution $p$, as shown in Fig. 2.6(b)-(d). To bring evidence to this observation, we performed a Monte Carlo study, where we drew one thousand sampling schemes, each one designed by solving the TSP on a set of independent random samples. We notice in Fig. 2.6 (d) that the empirical distribution of the points along the TSP curve, hereafter termed the final distribution, departs from the original distribution $p$. A simple intuition can be given to explain this discrepancy between the initial and final distributions in a $d$-dimensional acquisition space. Consider a small subset of the acquisition space $\omega$. In $\omega$, the number of points is proportional to $p$. The typical distance between two neighbors in $\omega$ is then proportional to $p^{-1/d}$. Therefore, the local length of the trajectory in $\omega$ is proportional to $pp^{-1/d} = p^{1-1/d} \neq p$. In what follows, we will show that the empirical measure of the TSP solution converges to a measure proportional to $p^{1-1/d}$.

### 2.4.2 Definitions

We shall work on the hypercube $\mathcal{H} = [0,1]^d$ with $d \geqslant 2$. In what follows, $\{x_i\}_{i\in\mathbb{N}^*}$ denotes a sequence of points in the hypercube $\mathcal{H}$, independently drawn from a density $p : \mathcal{H} \mapsto \mathbb{R}_+$. The set of the first $N$ points is denoted $X_N = \{x_i\}_{i\leqslant N}$.

Using the definitions introduced in Tab. 2.1, we introduce $\gamma_N : [0,1] \to \mathcal{H}$ the function that parameterizes $C(X_N)$ by moving along it at constant speed $T(X_N, \mathcal{H})$. Then, the *distribution of the TSP solution* reads as follows:

**Definition 2.16.** The distribution of the TSP solution is denoted $\tilde{P}_N$ and defined, for any Borelian $B$ in $\mathcal{H}$ by:

$$\tilde{P}_N(B) = \lambda_{[0,1]}\left(\gamma_N^{-1}(B)\right).$$

*Remark* 2.17. The distribution $\tilde{P}_N$ is defined for fixed $X_N$. It makes no reference to the stochastic component of $X_N$.

*Remark* 2.18. A more intuitive definition of $\tilde{P}_N$ can be given if we introduce other tools. For a subset $\omega \subseteq \mathcal{H}$, we denote the length of $C(X_N) \cap \omega$ as $T_{|\omega}(X_N, \mathcal{H}) = T(X_N, \mathcal{H})\tilde{P}_N(\omega)$. Using this definition, it follows that:

$$\tilde{P}_N(\omega) = \frac{T_{|\omega}(X_N, \mathcal{H})}{T(X_N, \mathcal{H})}, \ \forall \omega. \tag{2.9}$$

Then $\tilde{P}_N(\omega)$ is the relative length of the curve inside $\omega$.

### 2.4.3 Main results

Our main theoretical result introduced in (Chauffert et al., 2013a) reads as follows:

**Theorem 2.19.** *Define the density* $\tilde{p} = \dfrac{p^{(d-1)/d}}{\int_{\mathcal{H}} p^{(d-1)/d}(x)dx}$ *where* $p$ *is a density defined on* $\mathcal{H}$*. Then almost surely with respect to the law* $p^{\otimes\mathbb{N}}$ *of the random points sequence* $\{x_i\}_{i\in\mathbb{N}^*}$ *in* $\mathcal{H}$*, the distribution* $\tilde{P}_N$ *converges in distribution to* $\tilde{p}$*:*

$$\tilde{P}_N \xrightarrow{(d)} \tilde{p} \qquad\qquad p^{\otimes\mathbb{N}}\text{-}a.s. \tag{2.10}$$

The proof of the theorem is given in Appendix 4.

*Remark* 2.20. The $TSP$ solution does not define as such a VDS, since the underlying process is finite in time. Nevertheless, since $\tilde{P}_N$ is the occupation measure of $\gamma_N$, the following result holds:

*Corollary* 2.21. $(\gamma_N)_{N\in\mathbb{N}}$ *is a generalized* $\tilde{p}$ *VDS.*

*Remark* 2.22. The theorem indicates that if we want to reach distribution $p$ in 2D, we have to draw the initial points with respect to a distribution proportional to $p^2$, and to $p^{3/2}$ in 3D. Akin to the previous Monte Carlo study illustrating the behavior of the naive approach in Fig. 2.6 (top row), we repeated the same procedure after having taken this result into account. The results are presented in Fig. 2.6(e)-(g), in which it is shown that the final distribution now closely matches the original one (compare Fig. 2.6(g) with Fig. 2.6(a)).

*Remark* 2.23. Contrarily to the Markov chain approach for which we derived compressed sensing results in Proposition 2.11, the TSP approach proposed here is mostly heuristic and based on the idea that the TSP solution curve covers the space rapidly. An argument supporting this idea is the fact that in 2D, the TSP curve $C(X_N)$ does not self-intersect. This property is clearly lacking for random walks.

*Remark* 2.24. One of the drawback of this approach is the TSP's NP-hardness. We believe that this is not a real problem. Indeed, there now exist very efficient approximate solvers such as the Concorde solver (Applegate et al., 2006). It finds an approximate solution with $10^5$ cities from a few seconds to a few hours depending on the required accuracy of the solution. The computational time of the approximate solution is not a real limitation since the computation is done off-line from the acquisition procedure. Moreover, many solvers are actually designed in such a way that their solution also fulfil

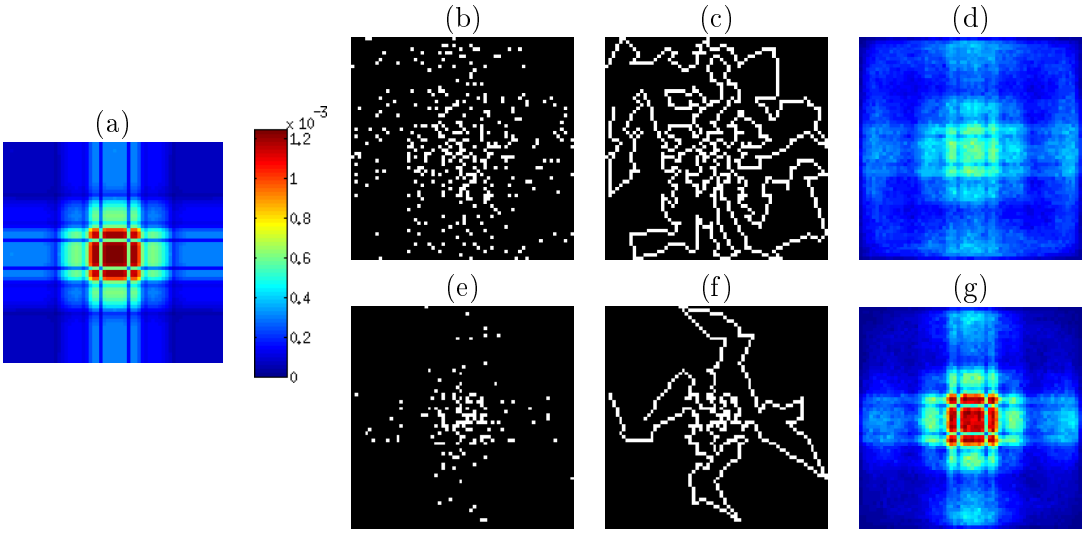


FIGURE 2.6: Illustration of the TSP-based sampling scheme to reach distribution $\pi$. (a): distribution $\pi$. (b) (resp. (e)): independent drawing of points from distribution $\pi$ (resp. $\propto \pi^2$). (c) (resp (f)): solution of the TSP amongst points of (b) (resp. (e)) . (d) and (g): Monte Carlo study: average scheme over one thousand drawings of sampling schemes, with the same color scale as in (a).

Theorem. 2.19. For example, in 2D, to reach a sampling factor of $R = 5$ on a $256 \times 256$ image, one need $N \simeq 10^4$ cities, and an approximate solution is obtained in 142s. In 3D, for a $256 \times 256 \times 256$ image, $N \simeq 9\ 10^5$ and an approximate solution is obtained in about 4 hours. In each case the solutions seem to be correctly approximated. In particular they do not self-intersect in 2D.

## 2.5 Experimental results in MRI

In this section, we focus on the reconstruction results by minimizing the $\ell_1$ problem (2.1) with a simple MRI model: $\mathbf{A} = \mathbf{F}^*\mathbf{\Psi}$, where $\mathbf{\Psi}$ denote the inverse Symmlet-10 transform[6]. The solution is computed using Douglas-Rachford's algorithm (Combettes and Pesquet, 2011). We consider an MR image of size $256 \times 256 \times 256$ as a reference, and perform reconstruction for different discrete sampling strategies. Every sampling scheme was regridded using a nearest neighbour approach to avoid data interpolation.[7]

[6]We focused on $\ell_1$ reconstruction since it is central in the CS theory. The reconstruction quality can be improved by considering more *a priori* knowledge on the image. Moreover we considered a simple MRI model, but our method can be extended to parallel MRI (Pruessmann et al., 1999), or spread-spectrum techniques (Haldar et al., 2011; Puy et al., 2012a).

[7]We provide Matlab codes to reproduce the proposed experiments here: http://chauffertn.free.fr/codes.html

### 2.5.1 2D-MRI

In 2D, we focused on a single slice of the MR image and considered its discrete Fourier transform as the set of possible measurements. First, we found the best made a comparison of independent drawings with respect to various distributions in order to find heuristically the best sampling density. Then we explored the performance of the two proposed methods to design continuous schemes: random walks and Travelling Salesman Problem. We also compared our solution to classical MRI sampling schemes. In every sampling schemes, the number of measurements is the same and equals 20% of the number of pixels in the image, so that the *sampling factor* $R$ is equal to 5. In cases where the sampling strategy is based on randomness (VDS, random walks, TSP...), we performed a Monte Carlo study by generating 100 sampling patterns for each variable density sampler.

#### 2.5.1.1 Variable density sampling using independent drawings

Here, we assessed the impact of changing the sampling distribution using independent drawings. In all experiments, we sampled the Fourier space center deterministically as shown on Figure 2.7.

TABLE 2.2: Quality of reconstruction results in terms of PSNR for 2D sampling with variable density independent drawings.

| | $\pi$ | polynomial decay: $(k_x^2 + k_y^2)^{-d/2}$ | | | | | |
|---|---|---|---|---|---|---|---|
| | | $d = 1$ | $\mathbf{d = 2}$ | $d = 3$ | $d = 4$ | $d = 5$ | $d = 6$ |
| mean PSNR (dB) | 35.6 | **36.4** | **36.4** | 36.3 | 36.0 | 35.5 | 35.2 |
| std dev. | $< 0.1$ | $< 0.1$ | $< 0.1$ | $< 0.1$ | $< 0.1$ | $< 0.1$ | $< 0.1$ |

Table 2.2 shows that the theoretically-driven optimal distribution $\pi$ is outperformed by the best heuristics. Amongst the latter, the distribution leading to the best reconstruction quality decays as $1/|k|^2$, which is the distribution used by Krahmer and Ward (Krahmer and Ward, 2014) as an approximation of $\pi$ for Haar wavelets. The standard deviation of the PSNR is negligible compared to the mean values and for a given distribution, each reconstrucion PSNR equals its average value at the precision used in Tab. 2.2.

#### 2.5.1.2 Continuous VDS

In this part we compared various variable density samplers:

- Random walks with a stationary distribution proportional to $1/|k|^2$ and different average chain lengths of $1/\alpha$,
- TSP-based sampling with distributions proportional to $1/|k|^2$ and $\pi$,
- Classical MRI sampling strategies such as spiral, radial and radial with random angles. The choice of the spiral follows Example 2.5: the spiral is parameterized by $s : [0, T] \to \mathbb{R}^2$, $\theta \mapsto r(\theta/T) \begin{pmatrix} \cos\theta \\ \sin\theta \end{pmatrix}$ where $r(t) := \frac{r(0)r(1)}{r(1)-t(r(1)-r(0))}$, so as the spiral density decays as $1/|k|^2$.

The sampling schemes are presented in Fig. 2.7 and the reconstruction results in Tab. 2.3.

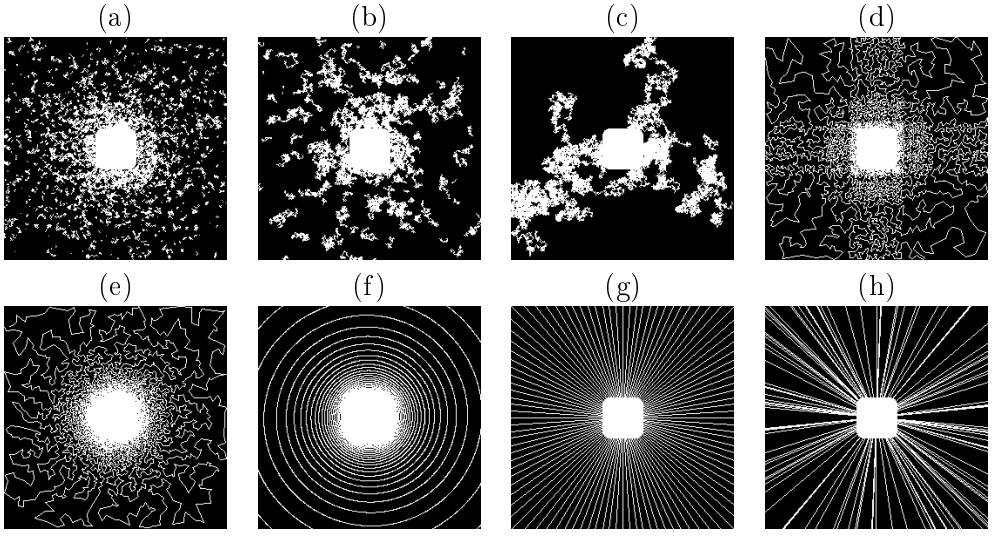


FIGURE 2.7: 2D continuous sampling schemes based on random walks with $\alpha = .1$ (a), $\alpha = .01$ (b), $\alpha = .001$ (c), and based on TSP solutions with distributions proportional to $\pi$ (d) and to $1/|k|^2$ (e). Classical sampling schemes: spiral (f), radial (g) and radial with random angles (h).

TABLE 2.3: Quality of reconstruction results in terms of PSNR for continuous sampling trajectories.

| | Markovian drawing ($\alpha$) | | | TSP sampling | | spiral | radial | radial random |
|---|---|---|---|---|---|---|---|---|
| | 0.1 | 0.01 | 0.001 | $\propto \pi$ | $\propto \mathbf{1/\lvert k\rvert^2}$ | | | |
| mean PSNR | 35.7 | 34.6 | 33.5 | 35.6 | **36.1** | 35.6 | 34.1 | 33.1 |
| std dev. | 0.1 | 0.3 | 0.6 | 0.1 | 0.1 | | | 0.4 |
| max value | 36.0 | 35.1 | 34.8 | 35.9 | 36.2 | | | 34.0 |
| in Fig. 2.7: | (a) | (b) | (c) | (d) | (e) | (f) | (g) | (h) |

As predicted by the theory, the shorter the chains the better the reconstructions. The optimal case corresponds to chains of length 1 ($\alpha = 1$) i.e. corresponding to independent VDS. When the chain is too long, large $k$-space areas are left unexplored, and the reconstruction quality decreases.

Besides, the use of a target distribution proportional to $1/|k|^2$ instead of $\pi$ for TSP-based schemes provides slightly better reconstruction results.

We also considered more classical sampling scheme. We observe that the spiral scheme and the proposed ones provide more accurate reconstruction results than radial schemes. We believe that the main reason underlying these different behaviors is closely related to the sampling rate decay from low to high frequencies, which is proportional to $1/|k|$ for radial schemes.

### 2.5.2 3D-MRI

Since VDS based on Markov chains have shown rather poor reconstruction results compared to the TSP-based sampling schemes in 2D simulations, we only focus on comparing TSP-based sampling schemes to classical CS sampling schemes. Moreover, the computational load to treat 3D images being much higher than in 2D, we only perform one drawing per sampling scheme in the following experiments. Experiments in 2D suggest that the reconstruction quality is not really impacted by the realization of a particular sampling scheme, except for drawing with Markov chains or with radial with random angles, which are not considered in our 3D experiments.

#### 2.5.2.1 Variable density sampling using independent drawings

The first step of the TSP-based approach is to identify a relevant target distribution. For doing so, we consider independent drawings as already done in 2D. The results are summarized in Tab. 2.4. In this experiment, we still use a number of measurements equal to 20% of the total amount ($R = 5$).

TABLE 2.4: Quality of reconstruction results in terms of PSNR for sampling schemes based on 3D variable density independent drawings, with densities $\propto 1/k^d$ and $\pi$, and with 20% of measured samples.

| $d$ | 1 | 2 | 3 | 4 | $\pi$ |
|---|---|---|---|---|---|
| PSNR (dB) | 44.78 | **45.01** | 44.56 | 44.03 | 42.94 |

The best reconstruction result is achieved with $d = 2$ and not the theoretically optimal distribution $\pi$. This illustrates the importance of departing from the sole sparsity hypothesis under which we constructed $\pi$. Natural signals have a much richer structure. For instance wavelet coefficients tend to become sparser as the resolution levels increase, and this feature should be accounted for to derive optimal sampling densities for natural images (see Section 2.6.)

#### 2.5.2.2 Efficiency of the TSP sampling based strategy

Let us now compare the reconstruction results using the TSP based method and the method proposed in the original CS-MRI paper (Lustig et al., 2007). These two sampling strategies are depicted in Fig. 2.8. For 2D independent drawings, we used the distribution providing the best reconstruction results in 2D, *i.e.* proportional to $1/|k|^2$. The TSP-based schemes were designed by drawing city locations independently with respect to a distribution proportional to $p^{\frac{3}{2}}$. According to Theorem 2.19 this is the correct way to reach distribution $p$ after joining the cities with constant speed along the TSP solution path. The experiments were performed with $p = \pi$ (see Fig. 2.2 (b)), and $p \propto 1/|k|^2$, since the latter yielded the best reconstruction results in the 3D independent VDS framework. We also compared these two continous schemes to 3D independent drawings with respect to a distribution proportional to $1/|k|^2$.

Reconstruction results with an sampling rate $R = 8.8$ are presented in Fig. 2.10, with a zoom on the cerebellum. The reconstruction quality using the proposed sampling scheme is better than the one obtained from classical CS acquisition and contains less artifacts. In particular, the branches of the cerebellum are observable with our proposed sampling scheme only. At higher sampling rate, we still observe less artifacts with the proposed schemes, as depicted in Fig. 2.11 with a sampling rate $R = 14.9$. Moreover, Fig. 2.9 shows that our proposed method outperforms the method proposed in (Lustig et al., 2007) by up to 2dB. If one aims at reaching a fixed PSNR, we can increase $r$ by more than 50% using the TSP based strategy. In other words, we could expect a substantial decrease of scanning time by using more advanced sampling strategies than those proposed until to now.

The two different choices of the target density $\pi$ and $\propto 1/|k|^2$ provide similar results. This is a bit surprising since 3D independent VDS with these two probability distributions provide very different reconstruction results (see Tab. 2.4). A potential explanation for that behavior is that the TSP tends to "smooth out" the target distribution. An independent drawing would collect very few Fourier coefficients in the blue zones of Fig. 2.2, notably the vertical and horizontal lines crossing the Fourier plane center.

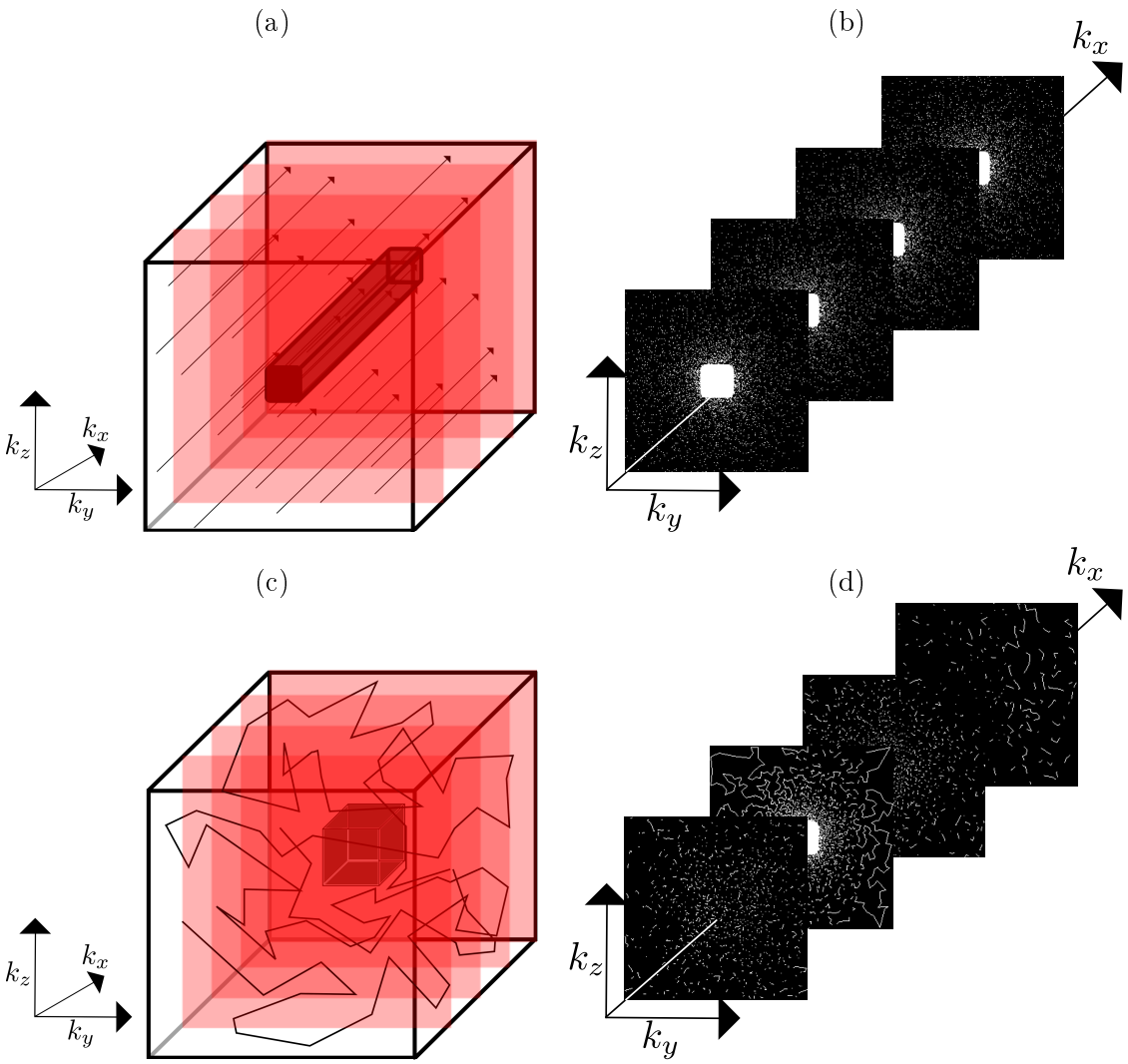


FIGURE 2.8: Compared sampling strategies in 3D-MRI. **Top:** 2D independent drawing sampling schemes designed by a planar independent drawing and measurements in the orthogonal readout direction. **Bottom:** 3D TSP-based sampling scheme. **Left:** Schematic representation of the 3D sampling scheme. **Right:** Representations of 4 parallel slices.

Sampling these zones seems to be of utmost importance since they contain high energy coefficients. The TSP approach tends to sample these zones by crossing the lines.

Perhaps the most interesting fact is that Fig. 2.9 shows that the TSP based sampling schemes provide results that are similar to independent drawings up to important sampling rates such as 20. We thus believe that the TSP solution proposed in this paper is near optimal since it provides results similar to unconstrained acquisition schemes. The price to be paid by integrating continuity constraints is thus almost null.

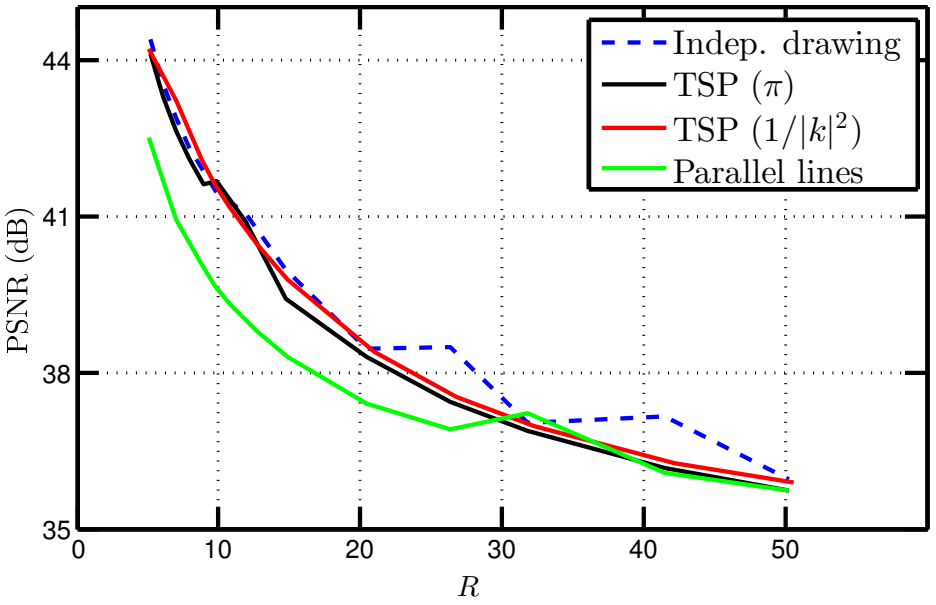


FIGURE 2.9: Quality of 3D reconstructed images in terms of PSNR as a function of sampling rates $R$ for various sampling strategies: independent drawings with respect to distribution $\propto 1/|k|^2$ (dashed blue line), TSP-based sampling with target densities $\pi$ (black line) and $\propto 1/|k|^2$ (red line), and parallel lines with 2D independent drawing with respect to $\propto 1/|k|^2$ distribution (green line) as depicted in Fig. 2.8[Top row].

## 2.6 Discussion and perspectives

In this paper, we investigated the use of variable density sampling along continuous trajectories. Our first contribution was to provide a well-grounded mathematical definition of $p$-variable density samplers (VDS) as stochastic processes with a prescribed limit empirical measure $p$. We identified through both theoretical and experimental results two key features characterizing their efficiency: their **empirical measure** as well as their **mixing properties**. We showed that VDS based on random walks were doomed to fail since they were unable to quickly cover the whole acquisition space. This led us to propose a two-step alternative that consists first of drawing random points independently and then joining them using a Travelling Salesman Problem solver. In contrast to what has been proposed in the literature so far, we paid attention to the manner the points have to be drawn so as to reach a prescribed empirical measure. Strikingly, our numerical results suggest that the proposed approach yields reconstruction results that are nearly equivalent to independent drawings. This suggests that adding continuity constraints to the sampling schemes might not be so harmful to derive CS results.

We believe that the proposed work opens many perspectives as outlined in what follows.

**How to select the target density?** We recalled existing theoretical results to address this point in Section 2.2 and showed that *deterministic* sampling could reduce the total

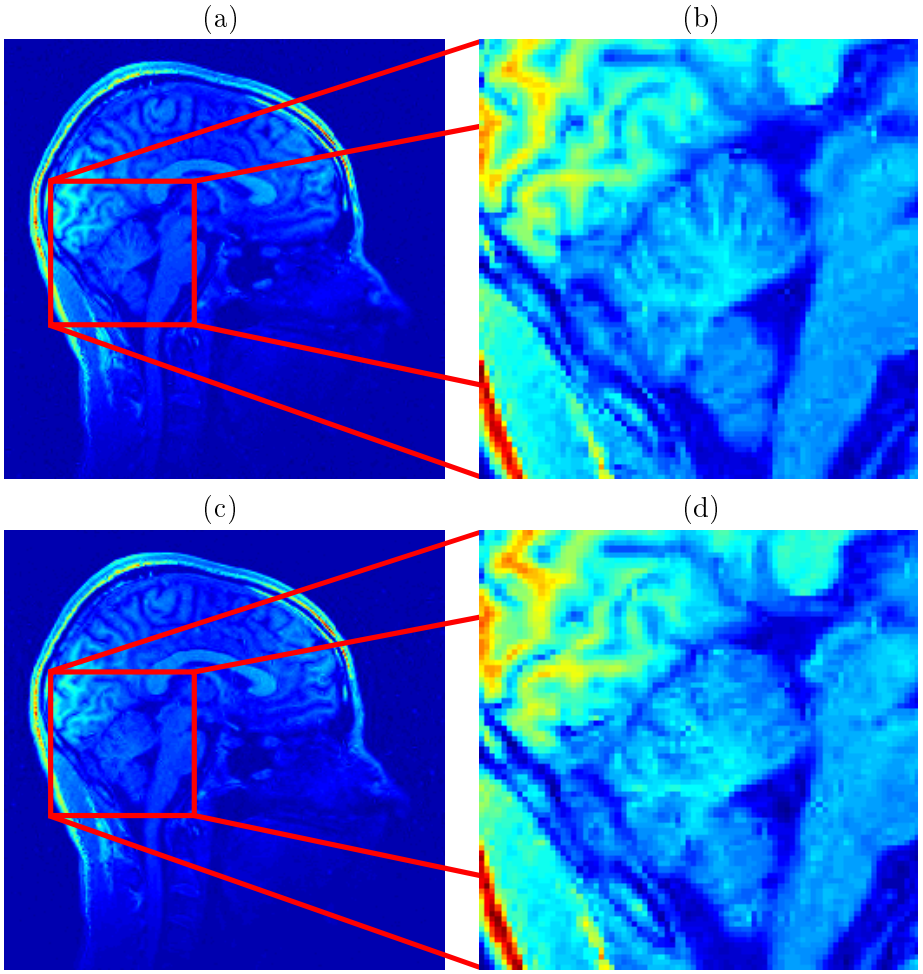


FIGURE 2.10: Reconstruction results for $R = 8.8$ for various sampling strategies. **Top row:** TSP-based sampling schemes (PSNR=42.1 dB). **Bottom row:** 2D random drawing and acquisitions along parallel lines (PSNR=40.1 dB). Sagital view (**left**) and zoom on the cerebellum (**right**).

number of required measurements. The analysis we performed closely followed the proofs proposed in (Rauhut, 2010; Candès and Plan, 2011) and was based solely on sparsity hypotheses on the signal/image to be reconstructed. The numerical experiments we performed indicate that heuristic densities still outperform the theoretical optimal ones. This suggests that the optimality critera used so far to derive target sampling densities does not account for the whole structure of the sought signal/image. Although sparsity is a key feature that characterizes natural signals/images, we believe that introducing stronger knowledge like *structured sparsity* might contribute to derive a new class of optimal densities that would compete with heuristic densities.

To the best of our knowledge, the recent paper (Adcock et al., 2013) is the first contribution that addresses the design of sampling schemes by accounting for a simple structured sparsity hypothesis. The latter assumes that wavelet coefficients become sparser as the

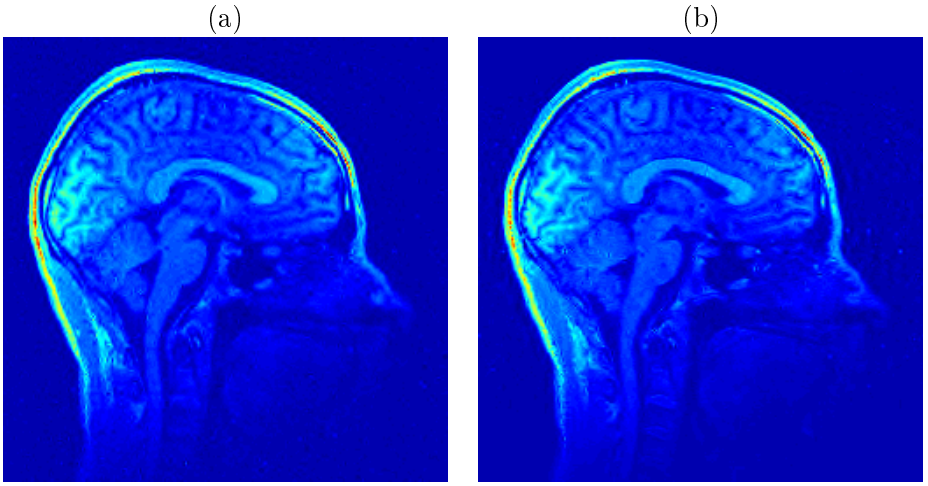


FIGURE 2.11: Reconstruction results for $R = 14.9$ for various sampling strategies. **Left:** TSP-based sampling schemes (PSNR=39.8 dB). **Right:** 2D random drawing and acquisitions along parallel lines (PSNR=38.3 dB).

resolution increases. The main conclusion of the authors is the same as that of Theorem 2.9 even though it is based on different arguments: the low frequencies of a signal should be sampled deterministically.

Finally, let us notice that the best empirical convex reconstruction techniques do not rely on the resolution of a simple $\ell^1$ problem such as (2.1). They are based on regularization with redundant frames and total variation for instance (Boyer et al., 2012). The signal model, the target density and the reconstruction algorithm should clearly be considered simultaneously to make a substantial leap on reconstruction guarantees.

**What VDS properties govern their practical efficiency?** In Section 2.3, it was shown that the key feature characterizing random walks efficiency was the mixing properties of the associated stochastic transition matrix. In order to derive CS results using generic random sets rather than point processes or random walks, it seems important to us to find an equivalent notion of mixing properties.

**How to generate VDS with higher degrees of regularity?** This is probably the most important question from a practical point of view. We showed that the TSP based VDS outperformed more conventional sampling strategies by substantial acceleration factors for a given PSNR value or recovers 3D images at an improved PSNR for a given acceleration factor. However, this approach may not really be appealing for many applications: continuity is actually not a sufficient condition for making acquisition sequences implementable on devices like MRI scanners or robot motion where additional kinematic constraints such as bounded first (gradients) and second (slew rate) derivatives should

be taken into account. Papers such as (Lustig et al., 2008) derive time-optimal waveforms to cross a given curve using optimal control. By using this approach, it can be shown that the angular points on the TSP trajectory have to be visited with a zero speed. This strongly impacts the scanning time and the distribution of the parametrized curve. The simplest strategy to reduce scanning time would thus consist in smoothing the TSP trajectory, however this approach dramatically changes the target distribution which was shown to be a key feature of the method. The key element to prove our TSP Theorem 2.19 was the famous Beardwood, Halton and Hammersley theorem (Beardwood et al., 1959). To the best of our knowledge, extending this result to smooth trajectories remains an open question[8]. Recent progresses in that direction were obtained in recent papers such as (Le Ny et al., 2012), but they do not provide sufficient guarantees to extend Theorem 2.19. Answering this question is beyond the scope of this paper. We believe that the work (Teuber et al., 2011) based on attraction and repulsion potentials opens an appealing research avenue for solving this issue.

## Appendix 1 - proof of Theorem 2.9

For a symmetric matrix $M$, we denote by $\lambda_{\max}(M)$ its largest eigenvalue and by $\|M\|$ the largest eigenvalue modulus. The crucial step to obtain Theorem 2.9 is Proposition 2.25 below. The rest of the proof is the same as the one proposed in (Rauhut, 2010) and we refer the interested reader to (Rauhut, 2010, Section 7.3) for further details.

**Proposition 2.25.** *Let $\Omega = \Omega_1 \cup \Omega_2 \subseteq \{1, \dots, n\}$ be a set constructed as in Theorem 2.9. Define*

$$\tilde{a}_i = \begin{cases} a_i & \text{if } i \in \Omega_1 \\ a_i/\sqrt{p_i} & \text{if } i \in \{1 \dots n\} \setminus \Omega_1. \end{cases}$$

*and*

$$\tilde{\mathbf{A}} = \begin{pmatrix} \tilde{a}_{\Omega_1(1)} \\ \vdots \\ \tilde{a}_{\Omega_1(m_1)} \\ \frac{1}{\sqrt{m_2}} \tilde{a}_{\Omega_2(1)} \\ \vdots \\ \frac{1}{\sqrt{m_2}} \tilde{a}_{\Omega_2(m_2)} \end{pmatrix} \in \mathbb{C}^{m \times n}. \tag{2.11}$$

[8] To be precise, many crucial properties of the length of the shortest path used to derive asymptotic results are lost. The most important one is subadditivity (Steele, 1981).

*Then for all* $\delta \in [0, \frac{1}{2}]$*:*

$$\mathbb{P}\Big(\Big\|\tilde{\mathbf{A}}^{S*}\tilde{\mathbf{A}}^{S} - I_s\Big\| \geqslant \delta\Big) \leqslant 2s \exp\left(-\frac{m_2\delta^2}{CK_2^2 s}\right).$$

*where* $\tilde{\mathbf{A}}^S \in \mathbb{C}^{m\times s}$ *is the matrix composed of the* $s$ *columns of* $\tilde{\mathbf{A}}$ *belonging to* $S$*.* $C = 7/3$ *is a constant.*

The proof of this proposition relies heavily on the matrix Bernstein inequality below (Tropp, 2012).

**Proposition 2.26** (Matrix Bernstein inequality)**.** *Let* $\mathbf{Z}_k$ *be a finite sequence of independent, random, self-adjoint matrices in* $\mathbb{C}^{d\times d}$*. Assume that each random matrix satisfies*

$$\mathbb{E}(\mathbf{Z}_k) = 0 \qquad \textit{and} \qquad \lambda_{\max}(\mathbf{Z}_k) \leqslant R \qquad \textit{a.s.}$$

*Denote* $\sigma^2 = \Big\|\sum_k \mathbb{E}(\mathbf{Z}_k^2)\Big\|$*. Then, for all* $t \geqslant 0$*:*

$$\mathbb{P}\Big(\Big\|\sum_k \mathbf{Z}_k\Big\| \geqslant t\Big) \leqslant 2d \exp\Big(-\frac{t^2/2}{\sigma^2 + Rt/3}\Big).$$

We are now ready to prove Proposition 2.25.

*Proof.* For any vector $v \in \mathbb{C}^n$, denote by $v^S \in \mathbb{C}^s$ the vector composed of the entries of $v$ belonging to $S \subseteq \{1,\dots,n\}$. Consider the random sequence $X_1,\dots,X_{m_2}$ where $X_i = j \in \{1\dots n\} \setminus \Omega_1$ with probability $p_j$, and denote by $\Omega_2$ the set $\{X_1,\dots X_{m_2}\}$. Denote by $\mathbf{M}_1 := \sum_{i\in\Omega_1} a_i^S a_i^{S*}$. Consider the matrices $\mathbf{Z}_j := \mathbf{M}_1 + \tilde{a}_j^S \tilde{a}_j^{S*} - I_s$. According to Eq. (2.11), we get by construction:

$$\tilde{\mathbf{A}}^{S*}\tilde{\mathbf{A}}^{S} - I_s = \frac{1}{m_2}\sum_{j\in\Omega_2} \mathbf{Z}_j.$$

Since $I_s = \sum_{i=1}^n a_i^S a_i^{S*}$, we notice that $\forall i \in \{1,\dots,m_2\}$ (i) $\mathbb{E}(\mathbf{Z}_{X_i}) = 0$, (ii) $\mathbb{E}(\tilde{a}_{X_i}^S \tilde{a}_{X_i}^{S*}) = I_s - \mathbf{M}_1$. Moreover, we have (iii) $0 \preceq I_s - \mathbf{M}_1 \preceq I_s$ and (iv) $0 \preceq \mathbf{M}_1 \preceq I_s$.

Using the identity $(\tilde{a}_j^S \tilde{a}_j^{S*})^2 = \|\tilde{a}_j^S\|^2 \tilde{a}_j^S \tilde{a}_j^{S*}$ and the fact that $\|\tilde{a}_i^S\| \leqslant \sqrt{s}\|\tilde{a}_i^S\|_\infty$, we get $\mathbb{E}((\tilde{a}_{X_i}^S \tilde{a}_{X_i}^{S*})^2) \preceq K_2^2 s(I_s - \mathbf{M}_1)$ using (ii). We can then proceed as follows using points

(iii) to (iv):

$$\begin{aligned}
\mathbb{E}(\mathbf{Z}_{X_i}^2) &= \mathbf{M}_1^2 - 2\mathbf{M}_1 + I_s + \mathbb{E}((\tilde{a}_{X_i}^S \tilde{a}_{X_i}^{S*})^2) + 2\mathbf{M}_1 \mathbb{E}(\tilde{a}_{X_i}^S \tilde{a}_{X_i}^{S*}) - 2\mathbb{E}(\tilde{a}_{X_i}^S \tilde{a}_{X_i}^{S*}) \\
&\leq \mathbf{M}_1^2 - 2\mathbf{M}_1 + I_s + K_2^2 s(I_s - \mathbf{M}_1) + 2\mathbf{M}_1(I_s - \mathbf{M}_1) - 2(I_s - \mathbf{M}_1) \\
&= -(I_s - \mathbf{M}_1)^2 + K_2^2 s(I_s - \mathbf{M}_1) \\
&\preceq K_2^2 s I_s.
\end{aligned}$$

Then $\| \sum_{i=1}^{m_2} \mathbb{E}(\mathbf{Z}_{X_i}^2) \| \leqslant m_2 K_2^2 s$.

By noticing that $\tilde{a}_{X_i}^S \tilde{a}_{X_i}^{S*} - I_s \preceq \mathbf{Z}_{X_i} \preceq \tilde{a}_{X_i}^S \tilde{a}_{X_i}^{S*}$, we obtain $\|\mathbf{Z}_{X_i}\| \leqslant K_2^2 s$. Finally, by applying Bernstein inequality to the sequence of matrices $\mathbf{Z}_{X_1}, \dots \mathbf{Z}_{X_{m_2}}$, we derive for all $t \geqslant 0$:

$$\mathbb{P}\Big( \Big\| \sum_{j \in \Omega_2} \mathbf{Z}_j \Big\| \geqslant t \Big) \leqslant 2s \exp\left( - \frac{t^2/2}{m_2 K_2^2 s + K_2^2 s t/3} \right).$$

Plugging $\delta := t/m_2$, and noticing that $\delta \leqslant 1/2 \Rightarrow 2(1+\delta/3) \leqslant 2(1+\delta/3) \leqslant 7/3$, the announced result is shown. □

## Appendix 2 - proof of Proposition 2.11

Our approach relies on the following perfect recovery condition introduced in (Juditsky and Nemirovski, 2011):

**Proposition 2.27** ((Juditsky and Nemirovski, 2011)). *If $\mathbf{A}_\Omega \in \mathbb{R}^{m \times n}$ satisfies*

$$\gamma(\mathbf{A}_\Omega) = \min_{\mathbf{Y} \in \mathbb{R}^{m \times n}} \|I_n - \mathbf{Y}^T \mathbf{A}_\Omega\|_\infty < \frac{1}{2s},$$

*all $s$-sparse signals $x \in \mathbb{R}^n$ are recovered exactly by solving the $\ell_1$ minimization problem (2.1).*

We noted $\|A\|_\infty$ the maximal modulus of all the entries of $A$. This can be seen as an alternative to the *mutual coherence* (Donoho, 2006). We limit our proof to the real case but it could be extended to the complex case using a slightly different proof.

We aim at finding $\mathbf{Y} \in \mathbb{R}^{m \times n}$, such that $\|I_n - \mathbf{Y}^T \mathbf{A}_\Omega\|_\infty < \frac{1}{2s}$, for a given positive integer $s$, where $\mathbf{A}_\Omega$ is the sensing matrix defined in Proposition 2.11. Following (Juditsky et al., 2011), we set $\mathbf{\Theta}_i = \frac{a_i a_i^T}{p_i}$ and use the decomposition $I_n = \mathbf{A}^T \mathbf{A} = \sum_{i=1}^n p_i \mathbf{\Theta}_i$. We consider a realization of the Markov chain $X_1, \dots, X_m$ , with $X_1 \sim p$ and $X_i \sim \mathbf{P}_{X_{i-1},:}$ for $i > 1$. Let us denote $\mathbf{W}_m = \frac{1}{m} \sum_{l=1}^m \mathbf{\Theta}_{X_l}$. Then $\mathbf{W}_m$ may be written as $\mathbf{Y}^T \mathbf{A}_\Omega$.

**Lemma 2.28.** $\forall\ 0 < t \leqslant 1$,

$$\mathbb{P}\left(\|I_n - \mathbf{W}_m\|_\infty \geqslant t\right) \leqslant n(n+1)e^{\frac{\epsilon(\mathbf{P})}{5}} \exp\Big(-\frac{mt^2\epsilon(\mathbf{P})}{12K^2(\mathbf{A},p)}\Big). \tag{2.12}$$

Before proving the lemma, let first recall a concentration inequality for finite-state Markov chains (Lezaud, 1998).

**Proposition 2.29.** *Let $(\mathbf{P}, p)$ be an irreducible and reversible Markov chain on a finite set $G$ of size $n$ with transition matrix $\mathbf{P}$ and stationary distribution $p$. Let $f : G \to \mathbb{R}$ be such that $\sum_{i=1}^n p_i f_i = 0$, $\|f\|_\infty \leqslant 1$ and $0 < \sum_{i=1}^n f_i^2 p_i \leqslant b^2$. Then, for any initial distribution $q$, any positive integer $m$ and all $0 < t \leqslant 1$,*

$$\mathbb{P}\Big(\frac{1}{m}\sum_{i=1}^m f(X_i) \geqslant t\Big) \leqslant e^{\frac{\epsilon(\mathbf{P})}{5}} N_q \exp\Big(-\frac{mt^2\epsilon(\mathbf{P})}{4b^2(1+g(5t/b^2))}\Big)$$

*where $N_q = (\sum_{i=1}^n (\frac{q_i}{p_i})^2 p_i)^{1/2}$ and $g$ is given by $g(x) = \frac{1}{2}(\sqrt{1+x} - (1 - x/2))$.*

Now, we can prove Lemma 2.28

*Proof.* By applying Proposition 2.29 to a function $f$ and then to its opposite $-f$, we get:

$$\mathbb{P}\Big(\Big|\frac{1}{m}\sum_{i=1}^m f(X_i)\Big| \geqslant t\Big) \leqslant 2e^{\frac{\epsilon(\mathbf{P})}{5}} N_q \exp\Big(-\frac{mt^2\epsilon(\mathbf{P})}{4b^2(1+g(5t/b^2))}\Big).$$

Then we set $f(X_i) = (I_n - \mathbf{\Theta}_{X_i})^{(a,b)}/K(\mathbf{A},p)$ as real-valued function. Recall that $p$ satisfies $\sum_{i=1}^n p_i f(X_i) = 0$. Since $\|f\|_\infty \leqslant 1$, $b = 1$ and $t \leqslant 1$, we deduce $1 + g(5t) < 3$. Moreover, since the initial distribution is $p$, $q_i = p_i, \forall i$ and thus $N_q = 1$. Finally, resorting to a union bound enables us to extend our result for the $(a,b)$th entry to the whole infinite norm of the $n \times n$ matrix $I_n - \mathbf{W}_m$ (2.12). □

Finally, set $s \in \mathbb{N}^*$ and $\eta \in (0,1)$. If $m$ satisfies Ineq. (2.8), then

$$\mathbb{P}\left(\|I_n - \mathbf{W}_m\|_\infty \geqslant \frac{1}{2s}\right) < \eta\,.$$

In other words, with probability at least $1-\eta$, every $s$-sparse signal can be recovered by $\ell_1$ minimization (2.1).

*Remark* 2.30. It is straightforward to derive a similar result to Theorem 2.9 and thus to justify that a partial deterministic sampling reduces the total number of measurements required for perfect recovery.

## Appendix 3 - proof of Remark 2.15

In this part, we prove that for a random walk with uniform stationary distribution $p$, $\epsilon(\mathbf{P}) = O(n^{-\frac{1}{d}})$. We use geometric bounds known as *Cheeger's inequality* in (Diaconis and Stroock, 1991) and *Conductance Bounds* in (Jerrum and Sinclair, 1989; Brémaud, 1999). Let us recall a useful result concerning finite state space irreducible reversible transition matrices $\mathbf{P}$.

The *capacity* of a set $B \subset \{1, \dots, n\}$ is defined as $p(B) := \sum_{i \in B} p(i)$ and the *ergodic flow* out of $B$ is defined by $F(B) := \sum_{i \in B, j \in B^c} p(i)\mathbf{P}_{i,j}$. The *conductance* of the pair $(\mathbf{P}, p)$ is:

$$\varphi(\mathbf{P}) := \inf_B \left( \frac{F(B)}{p(B)};\ \ 0 < |B| < n,\ \ p(B) \leqslant \frac{1}{2} \right).$$

Then the following result holds (see (Jerrum and Sinclair, 1989) and (Brémaud, 1999, Theorem 4.3)):

**Proposition 2.31.**

$$\frac{\varphi(\mathbf{P})^2}{2} \leqslant \epsilon(\mathbf{P}) \leqslant 2\varphi(\mathbf{P}).$$

Now, assume that $n^{1/d} \in \mathbb{N}$ is even and construct a finite graph with $n$ nodes representing a Euclidean grid of the unit hypercube of dimension $d$. Assume that the vertices of the graph at one grid point are the $2d$ nearest nodes, with periodic boundary conditions (the graph can be seen as a $d$-dimensional torus). Assume that the transition probability is uniform over the neighbors, thus the stationary distribution is also the uniform one. This graph is depicted in Fig. 2.12[Left], with $d = 2$.

Let $B$ be the halved graph defined by the hyperplane parallel to an axis of the grid and including its center, so that $p(B) = 1/2$. An illustration is given in 2D in Fig. 2.12[Right]. Since we assumed periodic boundary conditions, the number of nodes belonging to $B$ and having a neighbor in $B^c$ is $2n^{(d-1)/d}$. Each of these nodes have $2d$ neighbors, but only one belonging to $B^c$. Since the stationary distribution is equal to $1/n$ on each node, the ergodic flow is $2n^{(d-1)/d}(\frac{1}{n}\frac{1}{2d})$. It follows that $\epsilon(\mathbf{P}) \leqslant \frac{4}{d} n^{-\frac{1}{d}}$.

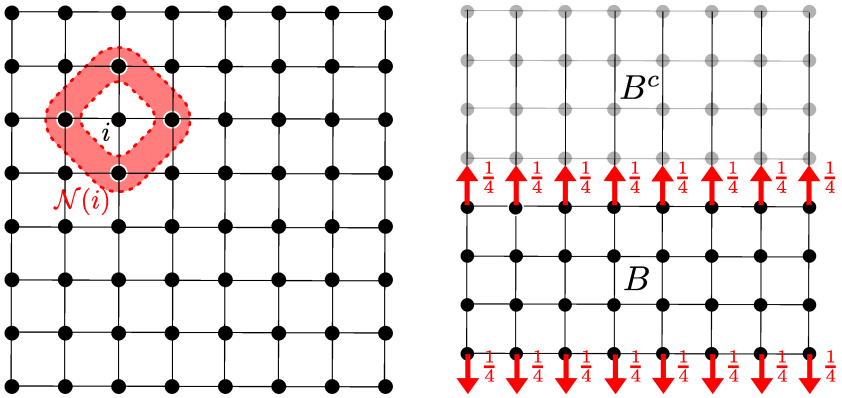


FIGURE 2.12: Illustration of the proof of Remark 2.15 in dimension 2. **Left:** regular grid with $n = \sqrt{n} \times \sqrt{n}$ nodes. **Right:** Graph partitioning in $B$ and $B^c$ with $p(B) = 1/2$.

## Appendix 4 - proof of Theorem 2.19

Let $h \in \mathbb{N}$. The set $\mathcal{H} = [0,1]^d$ will be partitioned in $h^d$ congruent hypercubes $(\omega_i)_{i \in I}$ of edge length $1/h$. The following proposition is central to obtain the proof:

**Proposition 2.32.** *Almost surely, for all $\omega_i$ in $\{\omega_i\}_{1 \leq i \leq h^d}$:*

$$\lim_{N \to \infty} \tilde{P}_N(\omega_i) = \tilde{p}(\omega_i) \tag{2.13}$$

$$= \frac{\int_{\omega_i} p^{(d-1)/d}(x)\mathrm{d}x}{\int_{\mathcal{H}} p^{(d-1)/d}(x)\mathrm{d}x} \qquad p^{\otimes \mathbb{N}}\text{-}a.s. \tag{2.14}$$

The strategy consists in proving that $T_{|\omega_i}(X_N, \mathcal{H})$ tends asymptotically to $T(X_N, \omega_i)$. The estimation of each term can then be obtained by applying the asymptotic result of Beardwood, Halton and Hammersley (Beardwood et al., 1959; Steele, 1981):

**Theorem 2.33.** *If $R$ is a Lebesgue-measurable set in $\mathbb{R}^d$ such that the boundary $\partial R$ has zero measure, and $\{y_i\}_{i \in \mathbb{N}^*}$, with $Y_N = \{y_i\}_{i \leq N}$ is a sequence of i.i.d. points from a density $p$ supported on $R$, then, almost surely,*

$$\lim_{N \to \infty} \frac{T(Y_N, R)}{N^{(d-1)/d}} = \beta(d) \int_R p^{(d-1)/d}(x)\mathrm{d}x, \tag{2.15}$$

*where $\beta(d)$ depends on the dimension $d$ only.*

To show Prop.2.32, we need to introduce the boundary TSP. For a set of points $F$ and an area $R$, we denote by $T_B(F, R)$ its length on the set $F \cap R$. The boundary TSP is defined as the shortest Hamiltonian tour on $F \cap R$ for the metric obtained from the Euclidean

metric by the quotient of the boundary of $R$, that is $d(a,b) = 0$ if $a, b \in \partial R$. Informally, it matches the original TSP while being allowed to travel along the boundary for free. We refer to (Frieze and Yukich, 2002) for a complete description of this concept.

We shall use a set of classical results on TSP and boundary TSP, that may be found in the survey books (Frieze and Yukich, 2002) and (Yukich, 1998).

**Lemma 2.34.** *Let $F$ denote a set of $n$ points in $\mathcal{H}$.*

1. *The boundary TSP is superadditive, that is, if $R_1$ and $R_2$ have disjoint interiors.*

$$T_B(F, R_1 \cup R_2) \geqslant T_B(F, R_1) + T_B(F, R_2). \tag{2.16}$$

2. *The boundary TSP is a lower bound on the TSP, both globally and on subsets. If $R_2 \subset R_1$:*

$$T(F, R) \geqslant T_B(F, R) \tag{2.17}$$

$$T_{|R_2}(F, R_1) \geqslant T_B(F, R_2) \tag{2.18}$$

3. *The boundary TSP approximates well the TSP (Yukich, 1998, Lemma 3.7)):*

$$|T(F, \mathcal{H}) - T_B(F, \mathcal{H})| = O(n^{(d-2)/(d-1)}). \tag{2.19}$$

4. *The TSP in $\mathcal{H}$ is well-approximated by the sum of TSPs in a grid of $h^d$ congruent hypercubes (Frieze and Yukich, 2002, Eq. (33)).*

$$|T(F, \mathcal{H}) - \sum_{i=1}^{h^d} T(F, \omega_i)| = O(n^{(d-2)/(d-1)}). \tag{2.20}$$

We now have all the ingredients to prove the main results.

*Proof of Proposition 2.32.*

$$\begin{aligned}\sum_{i\in I} T_B(X_N,\omega_i) &\overset{(2.16)}{\leqslant} T_B(X_N,\mathcal{H})\\ &\overset{(2.17)}{\leqslant} T(X_N,\mathcal{H}) = \sum_{i=1}^{h^d} T|_{\omega_i}(X_N,\mathcal{H})\\ &\overset{(2.20)}{\leqslant} \sum_{i=1}^{h^d} T(X_N,\omega_i) + O(N^{(d-1)/(d-2)})\end{aligned}$$

Let $N_i$ be the number of points of $X_N$ in $\omega_i$.

Since $N_i \leqslant N$, we may use the bound (2.19) to get:

$$\lim_{N\to\infty} \frac{T(X_N,\omega_i)}{N^{(d-1)/d}} = \lim_{N\to\infty} \frac{T_B(X_N,\omega_i)}{N^{(d-1)/d}}. \tag{2.21}$$

Using the fact that there are only finitely many $\omega_i$, the following equalities hold almost surely:

$$\begin{aligned}\lim_{N\to\infty} \frac{\sum_{i=1}^{h^d} T_B(X_N,\omega_i)}{N^{(d-1)/d}} &= \lim_{N\to\infty} \frac{\sum_{i=1}^{h^d} T(X_N,\omega_i)}{N^{(d-1)/d}}\\ &\overset{(2.20)}{=} \lim_{N\to\infty} \frac{\sum_{i=1}^{h^d} T_{|\omega_i}(X_N,\mathcal{H})}{N^{(d-1)/d}}.\end{aligned}$$

Since the boundary TSP is a lower bound (cf. Eqs. (2.18)-(2.17)) to both local and global TSPs, the above equality ensures that:

$$\begin{aligned}\lim_{N\to\infty} \frac{T_B(X_N,\omega_i)}{N^{(d-1)/d}} &= \lim_{N\to\infty} \frac{T(X_N,\omega_i)}{N^{(d-1)/d}}\\ &= \lim_{N\to\infty} \frac{T_{|\omega_i}(X_N,\mathcal{H})}{N^{(d-1)/d}} \qquad p^{\otimes\mathbb{N}}\text{-a.s, } \forall i.\end{aligned} \tag{2.22}$$

Finally, by the law of large numbers, almost surely $N_i/N \to p(\omega_i) = \int_{\omega_i} p(x)dx$. The law of any point $x_j$ conditioned on being in $\omega_i$ has density $p/p(\omega_i)$. By applying Theorem 2.33 to the hypercubes $\omega_i$ and $\mathcal{H}$ we thus get:

$$\lim_{N\to+\infty} \frac{T(X_N,\omega_i)}{N^{(d-1)/d}} = \beta(d)\int_{\omega_i} p(x)^{(d-1)/d}dx \qquad p^{\otimes\mathbb{N}}\text{-a.s, } \forall i.$$

and

$$\lim_{N\to+\infty} \frac{T(X_N,\mathcal{H})}{N^{(d-1)/d}} = \beta(d)\int_{\mathcal{H}} p(x)^{(d-1)/d}dx \qquad p^{\otimes\mathbb{N}}\text{-a.s, } \forall i.$$

Combining this result with Eqs. (2.22) and (2.9) yields Proposition 2.32. □

*Proof of Theorem 2.19.* Let $\varepsilon > 0$ and $h$ be an integer such that $\sqrt{d}h^{-d} < \varepsilon$. Then any two points in $\omega_i$ are at distance less than $\varepsilon$.

Using Theorem 2.32 and the fact that there is a finite number of $\omega_i$, almost surely, we get:

$\lim_{N\to+\infty} \sum_{i=1}^{h^d} \left|\tilde{P}_N(\omega_i) - \tilde{p}(\omega_i)\right| = 0$. Hence, for any $N$ large enough, there is a coupling $K$ of $\tilde{P}_N$ and $\tilde{p}$ such that both corresponding random variables are in the same $\omega_i$ with probability $1-\varepsilon$. Let $A \subseteq \mathcal{H}$ be a Borelian. The coupling satisfies $\tilde{P}_N(A) = K(A \otimes \mathcal{H})$ and $\tilde{p}(A) = K(\mathcal{H} \otimes A)$. Define the $\varepsilon$-neighborhood by $A^\varepsilon = \{X \in \mathcal{H} \,|\, \exists Y \in A,\ \|X - Y\| < \varepsilon\}$. Then, we have: $\tilde{P}_N(A) = K(A \otimes \mathcal{H}) = K(\{A \otimes \mathcal{H}\} \cap \{|X - Y| < \varepsilon\}) + K(\{A \otimes \mathcal{H}\} \cap \{|X - Y| \geqslant \varepsilon\})$. It follows that:

$$\begin{aligned}\tilde{P}_N(A) &\leqslant K(A \otimes A^\epsilon) + K(|X - Y| \geqslant \varepsilon) \\ &\leqslant K(\mathcal{H} \otimes A^\varepsilon) + \varepsilon = \tilde{p}(A^\varepsilon) + \varepsilon.\end{aligned}$$

This exactly matches the definition of convergence in the Prokhorov metric, which implies convergence in distribution. □

## Acknowledgments

The authors wish to thanks Yves Wiaux, Fabrice Gamboa, Jérémie Bigot, Laurent Miclo, Alexandre Vignaud and Claire Boyer for fruitful discussions and feedback. This research was supported by the Labex CIMI through a 3 months invitation of Philippe Ciuciu. This work was partially supported by ANR SPH-IM-3D (ANR-12-BSV5-0008), by the FMJH Program Gaspard Monge in optimization and operation research (MAORI project), and by the support to this program from EDF.

# Chapter 3

# A new concentration inequality

In this chapter, we introduce a new concentration inequality on the largest eigenvalue of the sum of zero-mean Hermitian matrix-valued Markov chain. This result generalizes the existing scalar concentration inequalities for Markov chains, as well as the case of sum of independent matrices, the dependencies being encoded by the spectral gap of the chain. We give an application in Compressed Sensing when the sampling is based on random walks that improves the proposition 2.11 page 42.

## 3.1 Introduction

We consider a finite graph $\mathcal{G}$ with $N$ vertices and an irreducible and reversible Markov chain $(X_n)$ on this graph. Let $f$ be a matrix-valued mapping from $\mathcal{G}$ to $\mathbb{H}^d$, the set of Hermitian matrices of size $d \times d$. If $\pi$ denotes the stationary distribution of the chain, we expect that $n^{-1}\sum_{i=1}^{n} f(X_i)$ converges to $\pi f = \sum_{y\in\mathcal{G}} \pi(y)f(y)$. Here, we are interested in controlling the deviation from the mean, i.e., the largest eigenvalue $\lambda_{\max}$ of the partial sum process in the case where $\pi f = 0$. Our concentration result reads:

**Proposition 3.1** (Bennett's inequality for matrix-valued Markov chains)**.** *Let $X_1, \ldots, X_n$ be an irreducible and reversible Markov chain and denote by $P$ its transition matrix with stationary distribution $\pi$ and spectral gap $\varepsilon$. Assume that $X_1 \sim q$ and that:*

$$\sum_{y\in\mathcal{G}} \pi(y)f(y) = 0 \quad \textit{and} \quad \lambda_{\max}(f(y)) \leqslant R, \forall y \in \mathcal{G}. \tag{3.1}$$

*Define the variance parameters:*

$$\sigma^2 := \lambda_{\max}\left(\sum_{y\in\mathcal{G}} \pi(y)f(y)^2\right), \quad \textit{and} \quad \sigma_n^2 = n\cdot\sigma^2$$

*Then, for all* $t > 0$,

$$\mathbb{P}\left(\lambda_{\max}\left(\sum_{i=1}^{n} f(X_i)\right) \geqslant t\right) \leqslant d \cdot \sup(\frac{q_i}{\pi_i}) \cdot \exp\left(\frac{2\sigma_n^2}{R^2\varepsilon} \cdot h\left(\frac{R \cdot \varepsilon t}{2\sigma_n^2}\right)\right), \tag{3.2}$$

*where* $h(x) = (1+x)\log(1+x)$.

A more common inequality is Bernstein's one. It is a direct consequence of Proposition 3.1.

**Proposition 3.2** (Bernstein's inequality for matrix-valued Markov chains)**.** *With the same hypothesis as in Prop 3.1, for all* $t > 0$*:*

$$\mathbb{P}\left(\lambda_{\max}\left(\sum_{i=1}^{n} f(X_i)\right) \geqslant t\right) \leqslant d \cdot \sup(\frac{q_i}{\pi_i}) \cdot \exp\left(-\frac{\varepsilon t^2}{4\sigma_n^2 + 2Rt\varepsilon/3}\right). \tag{3.3}$$

To the best of our knowledge, there is no such concentration inequality in the literature. Existing results consider either a sum of scalar-valued (or vector-valued) variables defined over a Markov chain, or a sum of *independent* Hermitian matrices. The contribution of this work is thus a generalization of these inequalities to matrix-valued Markov chains. Let us start by discussing the accuracy of our result introduced in Proposition 3.2.

### 3.1.1 Existing concentration results on real-valued Markov chains

One of the first results of this type was proven by Lezaud (Lezaud, 1998). We notice that the quantities involved in his result were really similar to our results introduced in Propositions 3.1 and 3.2. For a real mapping $f : \mathcal{G} \to \mathbb{R}$, such that $|f(y)| \leqslant R \ \forall y \in \mathcal{G}$ and $\pi f = 0$, he showed that for all $t > 0$:

$$\mathbb{P}\Big(\sum_{i=1}^{n} f(X_i) \geqslant t\Big) \leqslant e^{\varepsilon/5} N_q \exp\Big(-\frac{\varepsilon t^2}{4nb^2(1+g(5t/nb^2R))}\Big),$$

where $g$ is given by $g(x) = \frac{1}{2}(\sqrt{1+x} - (1 - x/2))$, $N_q = \left(\sum_{y\in\mathcal{G}} \pi(y)\frac{q(y)^2}{\pi(y)^2}\right)^{1/2}$, $b^2 = \sum_{y\in\mathcal{G}} \pi(y) f(y)^2$, and $\varepsilon$ denotes the spectral gap of the chain. This concentration results is based on Kato's theory on perturbation operator (Kato, 1976). Lezaud improved the concentrations results obtained by Gillman (Gillman, 1993) and Dinwoodie (Dinwoodie, 1995) that also used tools from operator perturbation theory. Kargin (Kargin, 2007) uses Kato's theory to prove a concentration inequality for vector-valued Markov chains.

Using a different approach based on Marton coupling, Paulin (Paulin, 2012a) proved Hoeffding and Bernstein inequalities using the more intuitive notion of mixing time instead of spectral properties of $P$. Mixing time is defined as

$$t_{\text{mix}}(\eta) = \min\left\{t : \sup_{x\in\Omega} d_{TV}(P^t(x,\cdot),\pi) \leqslant \eta\right\},$$

where $d_{TV}$ is the total variation distance, and $P^t(x,\cdot)$ is the distribution of the chain starting from $x$ with transition matrix $P$ at time $t$. He proved also that the mixing time $t_{\text{mix}}(\eta)$ is an upper bound of $1/\varepsilon$ times a constant that depends on the precision of the mixing time. His Bernstein inequality reads:

$$\mathbb{P}\Big(\sum_{i=1}^{n} f(X_i) \geq t\Big) \leq \exp\Big(-\frac{nt^2/\tau'_{\text{min}}}{8V + 4\sqrt{2}t}\Big),$$

where $V = \mathbb{E}\left(\sum_{i=1}^{N} f(X_i)^2\right)$, and $\tau'_{\text{min}} \simeq t_{\text{mix}}(1/4)$. In (Paulin, 2012a), the author shows that the mixing time is the natural quantity governing the speed of convergence. The framework is more general since it encompasses continuous and discrete, such as homogeneous and non-homogeneous chains (in particular, the last formula is valid even if the chain is not homogeneous). Finally, let us mention that Joulin and Ollivier (Joulin et al., 2010) obtained concentration inequalities for Lipschtiz functions with bounds depending on the Ricci curvature of the chain. The Ricci curvature is smaller than the spectral gap for reversible chains, providing weaker concentration bounds (see also (Paulin, 2014)). One of the motivation of this method is that there is no dependency on the initial ditribution, i.e., one can start from any location of the state space.

### 3.1.2 Sum of independent Hermitian matrices.

First works on matrix-valued concentration inequalities relied on the matrix Laplace transform for matrices, introduced by Ahlswede and Winter (Ahlswede and Winter, 2002). They extended Bernstein's method developed for the sum of independent scalar variables to the set of Hermitian matrices. Their result was then used in (Oliveira, 2009; Tropp, 2012) in the following form:

$$\mathbb{P}\Big(\lambda_{\max}\big(\sum_k X_k\big) \geqslant t\Big) \leqslant \inf_{\theta>0}\Big\{e^{-\theta t}\mathbb{E}\operatorname{tr}\exp\big(\theta\sum_k X_k\big)\Big\}, \quad \forall t > 0.$$

where $X_k$ are Hermitian random matrices. The first impediment of this method is that the so-called moment generating function $\mathbb{E}\exp\big(\theta\sum_k X_k\big)$ is *not* equal to $\mathbb{E}\prod_k \exp\big(\theta X_k\big)$ as in the scalar case since matrices do not commute in general. An alternative is to use

Golden-Thompson inequality (Golden, 1965; Thompson, 1965) that states

$$\operatorname{tr}\exp\left(\theta(X_1+X_2)\right) \leqslant \operatorname{tr}\exp(\theta X_1)\exp(\theta X_2),$$

but this inequality cannot be generalized to more than two matrices. Therefore, in the original papers (Ahlswede and Winter, 2002; Vershynin, 2009; Gross, 2011), the Golden-Thompson inequality was used within iterative methods to obtain concentration inequalities.

Further developments (Tropp, 2012; Oliveira, 2009) used more powerful tools based on Hermitian matrix properties. For example, using Lieb's theorem (Lieb, 1973), Tropp (Tropp, 2012) obtained the following concentration inequality, for a sequence of i.i.d. random Hermitian matrices $(Y_k)_{1\leqslant k\leqslant n}$ such that $\mathbb{E}^{(\pi)}Y_k = 0$ and $\lambda_{\max}(Y_k) \leqslant R$ almost surely:

$$\mathbb{P}\left(\lambda_{\max}\left(\sum_{k=1}^{n} Y_k\right) \geqslant t\right) \leqslant d\cdot\exp\left(-\frac{t^2}{2\sigma_n^2 + 2Rt/3}\right). \tag{3.4}$$

where $\sigma_n^2 = \lambda_{\max}\left(\sum_k \mathbb{E}Y_k^2\right)$[1]. The inequality obtained by Oliveira (Oliveira, 2009) only differs by weaker constants. Note that $\sigma_n^2$ is the norm of the sum of variances, whereas original Ahlswede and Winter's method (Ahlswede and Winter, 2002; Vershynin, 2009; Gross, 2011) involved the sum of variance norms, that is in general larger (see the discussion in (Tropp, 2012, Section 1.1)). In our context, since $\mathbb{E}^{(\pi)}f(X_i) = \sigma^2$ does not depend on $i$, hence these two variance parameters are equal. In an independent framework (corresponding to $\varepsilon = 1$), our result reads:

$$\mathbb{P}\left(\lambda_{\max}\left(\sum_{i=1}^{n} f(X_i)\right) \geqslant t\right) \leqslant d\cdot\exp\left(-\frac{t^2}{4\sigma_n^2 + 2Rt/3}\right), \tag{3.5}$$

that is similar to the stat-of-art, except that the constants are different (our bound is slightly weaker because of the $4\sigma_n$ term, instead of $2\sigma_n$ in (3.4)).

Finally, let us mention that a new technique for proving concentration inequality has recently emerged, based on a matrix generalization of Stein's method of exchangeable pairs (Paulin, 2012b; Paulin et al., 2013; Mackey et al., 2014). A similar Bernstein concentration inequality is obtained with this method (Mackey et al., 2014, Corollary 5.2). In addition, this technique permits to introduce weak dependencies between random variables.

[1] Here $\lambda_{\max}$ is equivalent to the spectral norm, since the matrix is positive.

### 3.1.3 Chapter overview

In the next section, we introduce the background that will be used to prove propositions 3.1 and 3.2. We introduce the notation of the paper, such as technical tools about Kronecker product. We also recall crucial properties on Hermitian matrices and on Markov chains that are the two key concepts hereafter. Then, section 3.3 contains the proof of our main results (prop. 3.1 and 3.2).

## 3.2 Preliminaries

### 3.2.1 General notation

We denote by $(e_i)_{1\leqslant i\leqslant N}$ the canonical basis of $\mathbb{C}^N$, and by $(f_i)_{1\leqslant i\leqslant d}$ the canonical basis of $\mathbb{C}^d$. For vectors $q, \pi \in \mathbb{R}^N$, we denote by $Q = diag(q_1, \dots, q_N) = diag(q)$ and $\Pi = diag(\pi)$ the diagonal matrices in $\mathbb{R}^{N\times N}$. We define $\mathbf{1} \in \mathbb{R}^N$ the vector containing ones. We denote by $F \in \mathbb{R}^{N\cdot d\times N\cdot d}$ the Hermitian block-diagonal matrix $diag(f(g_1), \dots, f(g_N))$, where $g_1, \dots g_N$ are the vertices of $\mathcal{G}$. Let us recall some properties of the Kronecker product:

**Proposition 3.3.** *If $A \in \mathbb{C}^{m\times n}$ and $B \in \mathbb{C}^{p\times q}$, we denote $A \otimes B \in \mathbb{C}^{m\cdot p\times n\cdot q}$ the matrix:*

$$\begin{pmatrix} a_{11}B & \cdots & a_{1n}B \\ \vdots & \ddots & \vdots \\ a_{m1}B & \cdots & a_{mn}B \end{pmatrix}.$$

*The Kronecker product satisfies the following properties:*

1. *If $A, B, C, D$ are four matrices and if the sizes are compatible,*

$$(A \otimes B)(C \otimes D) = (AC) \otimes (BD)$$

2. $(A \otimes B)^* = A^* \otimes B^*$

3. *If $A \in \mathbb{R}^{p\times p}$ and $B \in \mathbb{R}^{q\times q}$ are two hermitian matrices, and if $(\lambda_1, \dots, \lambda_p)$ and $(\mu_1, \dots, \mu_q)$ are the eigenvalues (with multiplicity) of $A$ and $B$, $A \otimes B$ is also Hermitian and diagonalizable with eigenvalues $(\lambda_i \cdot \mu_j, 1 \leqslant i \leqslant p, 1 \leqslant j \leqslant q)$*

For notation compacity, we also denote:

- $\underline{P} = P \otimes I_d \in \mathbb{C}^{N\cdot d\times N\cdot d}$, and $\underline{\Pi} = \Pi \otimes I_d \in \mathbb{C}^{N\cdot d\times N\cdot d}$:

$$\underline{P} = \begin{pmatrix} P_{11}I_d & P_{12}I_d & \dots & P_{1N}I_d \\ P_{21}I_d & P_{22}I_d & \dots & P_{2N}I_d \\ \vdots & & \ddots & \\ P_{N1}I_d & & & P_{NN}I_d \end{pmatrix} \quad \underline{\Pi} = \begin{pmatrix} \pi_1 \cdot I_d & 0 & \dots & 0 \\ 0 & \pi_2 \cdot I_d & \dots & 0 \\ \vdots & & \ddots & \\ 0 & & & \pi_N \cdot I_d \end{pmatrix}$$

- $\underline{e}_k = e_k \otimes I_d \in \mathbb{C}^{N \cdot d \times d}$. We also denote $\underline{\mathbf{1}} = \mathbf{1} \otimes I_d \in \mathbb{C}^{N \cdot d \times d}$ and $\underline{\pi} = \pi \otimes I_d \in \mathbb{C}^{N \cdot d \times d}$.

Finally, we equip the complex Hilbert space $\mathbb{C}^{N \cdot d}$ with the Hermitian product $\langle \cdot, \cdot \rangle_{\underline{\Pi}}$ defined by $\langle \underline{x}, \underline{y} \rangle_{\underline{\Pi}} = \underline{x}^* \underline{\Pi} \underline{y}$ and the complex Hilbert space $\mathbb{C}^N$ with the Hermitian product $\langle \cdot, \cdot \rangle_{\Pi}$ defined by $\langle x, y \rangle_{\Pi} = x^* \Pi y$.

### 3.2.2 The set of Hermitian matrices.

We denote the real vector space of Hermitian matrices by $\mathbb{H}^d := \{X \in \mathbb{C}^{n \times n}, X^* = X\}$, where $X^*$ denote the conjugate transpose of $X$. The eigenvalues of a matrix $X \in \mathbb{H}^d$ are real and we denote by $\lambda_{\max}(X)$ (resp. $\lambda_{\min}(X)$) its largest (resp. smallest) eigenvalue. tr denote the trace operator of a matrix. For $X_1, X_2 \in \mathbb{H}^d$, we say that $X_1 \preceq X_2$ if the eigenvalues of $X_2 - X_1$ are positive.

First, let us recall the following lemma,

**Lemma 3.4.** *For all $A, B \in \mathbb{H}^d$ such that $A \preceq B$, $tr(A) \leqslant tr(B)$.*

The following lemma was proposed by Tropp (Tropp, 2012) to bound the exponential of a zero-mean random matrix in expectation:

**Lemma 3.5.** *$\forall M \in \mathbb{H}^d$ a random matrix, such that $\lambda_{\max}(M) \leqslant 1$, and $\mathbb{E}M = 0$:*

$$\mathbb{E}e^{\theta M} \preceq \exp\left(g(\theta)\mathbb{E}(M^2)\right) \quad \forall \theta > 0$$

*where $g(\theta) = (e^\theta - \theta - 1)$.*

This lemma is crucial to obtain sharp bounds in the final concentration inequality. For comparison, Oliveira (Oliveira, 2009) used the weaker bound : $e^M \preceq I_d + M + M^2$. This explains that he obtained a larger bound in his Bernstein concentration inequality.

We now recall properties that hold for any random matrices. Let $M$ be a measurable matrix. The definition of the expectation implies that

$$\mathrm{tr}\mathbb{E}M = \mathbb{E}\mathrm{tr}M \tag{3.6}$$

$$\mathbb{E}M^* = (\mathbb{E}M)^* \tag{3.7}$$

### 3.2.3 Background on Markov chains

$\mathbb{E}^{(\pi)}$ will denote the expectation relative to the invariant distribution, whereas $\mathbb{E}^{(\pi)}$ will denote the expectation with respect to the initial distribution. We denote by $\mathbb{E}_i$ the conditional expectation $\mathbb{E}[\cdot|X_i]$.

Since $P$ defines a reversible Markov chain (i.e., such that $\Pi P = P^T\Pi$), its eigenvalues are real and its eigenvectors are orthogonal for the scalar product $\langle\cdot,\cdot\rangle_\Pi$ defined in section 3.2.1 (since $\Pi^{\frac{1}{2}}P\Pi^{-\frac{1}{2}}$ is a symmetric matrix, see (Brémaud, 1999) for details). We order the eigenvalues as follows:

$$1 = \lambda_1 > \lambda_2 \geqslant \ldots \geqslant \lambda_N \geqslant -1. \tag{3.8}$$

The inequality $\lambda_1 > \lambda_2$ comes from the irreducibility of the chain. $\mathbf{1}$ spans the eigenspace associated with the eigenvalue 1. $\varepsilon := 1 - \lambda_2$ is the spectral gap of the chain.

## 3.3 Proof of propositions 3.1 and 3.2

### 3.3.1 The Laplace transform method

The matrix Laplace transform is crucial in many proofs of matrix valued concentration inequalities. It was introduced by Ahlswede and Winter (Ahlswede and Winter, 2002), and a different proof is provided in (Oliveira, 2010; Tropp, 2012).

**Lemma 3.6** (Matrix Laplace transform)**.** *For all random matrix $M \in \mathbb{H}^d$:*

$$\mathbb{P}\Big(\lambda_{\max}\big(M\big) \geqslant t\Big) \leqslant \inf_{\theta>0}\Big\{e^{-\theta t}\mathbb{E}tr\exp\big(\theta M\big)\Big\}, \quad \forall t > 0.$$

The term $\mathbb{E}^{(\pi)}\mathrm{tr}\exp\big(\theta M\big)$ is called matrix moment generating function since the moments of the random matrix $M$ can be obtained by successive derivations and evaluations in

zero. In our case, Lemma 3.6 ensures that:

$$\mathbb{P}\left(\lambda_{\max}\left(\sum_{i=1}^{n} f(X_i)\right) \geqslant t\right) \leqslant \inf_{\theta>0}\left\{e^{-\theta t}\mathbb{E}\operatorname{tr}\exp\big(\theta\sum_{i=1}^{n} f(X_i)\big)\right\}. \tag{3.9}$$

Now, the quantity to control is the matrix moment generating function (mgf):

$$\mathbb{E}\operatorname{tr}\exp\big(\theta\sum_{i=1}^{n} f(X_i)\big) \tag{3.10}$$

The outline of the proof is classical (Tropp, 2012; Oliveira, 2009). We first use the Ahlswede and Winter transform to express our problem as the bounding of the trace of an exponential matrix (the mgf in Eq. (3.10)). Then, we try to find a relationship between the expectations $\mathbb{E}^{(\pi)}$ and $\mathbb{E}$, that is to say, how the expectation will change if the initial value $X_0$ is drawn from law $\pi$ or law $q$. Following (Tropp, 2012), we bound the moment generating function with only the second-order moment by resorting to lemma 3.5. Finally, the main difficulty of the proof is to bound this second-order moment in a Markov chain framework. We propose an upper-bound of this term that relies on the spectral properties of the chain.

### 3.3.2 Influence of the initial distribution

In this section, we are looking for a relation between $\mathbb{E}^{(\pi)}$ and $\mathbb{E}$. By definition,

$$\mathbb{E}\operatorname{tr}\exp\big(\theta\sum_{i=1}^{n} f(X_i)\big) = \sum_{X_1,\ldots,X_n} \operatorname{tr}\exp\left(\theta\sum_{i=1}^{n} f(X_i)\right) q(X_1)\prod_{i=2}^{n} P(X_{i-1},X_i),$$

here the summation is taken over all the possible trajectories $X_1,\ldots,X_n$ of the chain. Hence, we set

$$\begin{aligned}&\sum_{X_1,\ldots,X_n} \operatorname{tr}\exp\left(\theta\sum_{i=1}^{n} f(X_i)\right) q(X_1)\prod_{i=2}^{n} P(X_{i-1},X_i)\\ &\leqslant \sup(\frac{q_i}{\pi_i})\sum_{X_1,\ldots,X_n} \operatorname{tr}\exp\left(\theta\sum_{i=1}^{n} f(X_i)\right)\pi(X_1)\prod_{i=2}^{n} P(X_{i-1},X_i)\end{aligned}$$

Then, it comes the following lemma:

**Lemma 3.7.**

$$\mathbb{E}tr\exp\Big(\theta\sum_{i=1}^{n}f(X_i)\Big)\leqslant \sup(\frac{q_i}{\pi_i})\mathbb{E}^{(\pi)}tr\exp\Big(\theta\sum_{i=1}^{n}f(X_i)\Big) \tag{3.11}$$

This bound involves $\sup(\frac{q_i}{\pi_i})$, that also appear in the final results 3.1 and 3.2. For comparison, in the context of sum of scalar valued Markov chains, Lezaud (Lezaud, 1998) introduced an $\ell_2$-norm of $q/\pi$, instead of this $\ell_\infty$-norm. In (Paulin, 2012a), the author succeeded in removing this dependency from the concentration inequality. Finally notice that this term appear through a log term when we want to compute confidence intervals.

### 3.3.3 Control of the mgf with the second-order moment

Thanks to the previous section, we have to control the mgf $\mathbb{E}^{(\pi)}\mathrm{tr}\exp\big(\theta\sum_{i=1}^{n}f(X_i)\big)$, which corresponds to the expectation of a Markov chain generated with $X_1\sim\pi$ and with transition matrix $P$.

By lemmas 3.4 and 3.5, and using that trace and $\mathbb{E}^{(\pi)}$ commute, we have, for all $\theta>0$:

$$\mathbb{E}^{(\pi)}\mathrm{tr}\exp(\theta\sum_{i=1}^{n}f(X_i))=\mathrm{tr}\mathbb{E}^{(\pi)}\exp(\theta\sum_{i=1}^{n}f(X_i))\leqslant \mathrm{tr}\exp\left(g(\theta)\mathbb{E}^{(\pi)}\left(\sum_{i=1}^{n}f(X_i)\right)^2\right). \tag{3.12}$$

### 3.3.4 Control of second-order moment

In what follows, the main novelty of our result is exposed. It is largely based on matrix manipulations. We aim at controlling the term $\mathbb{E}^{(\pi)}\left(\sum_{i=1}^{n}f(X_i)\right)^2$. First let us develop the square:

$$\mathbb{E}^{(\pi)}\left(\sum_{i=1}^{n}f(X_i)\right)^2=\sum_{i=1}^{n}\sum_{j=1}^{n}\mathbb{E}^{(\pi)}\left(f(X_i)f(X_j)\right) \tag{3.13}$$

Let us introduce two crucial lemmas, proven in Appendices 1 and 2.

**Lemma 3.8.** *For all* $1\leqslant i,j\leqslant n$,

$$\mathbb{E}^{(\pi)}\left(f(X_i)f(X_j)\right)=\underline{\mathbf{1}}^TF\underline{\Pi P}^{|j-i|}F\underline{\mathbf{1}}.$$

**Lemma 3.9.** *For all $k \geqslant 0$*

$$\underline{\mathbf{1}}^T F \underline{\Pi P}^k F \underline{\mathbf{1}} \preceq \lambda_2^k \cdot \sigma^2 \cdot I_d,$$

*where $\lambda_2$ is the second largest eigenvalue of $P$, and where*

$$\sigma_2 = \lambda_{\max} \left( \sum_{y \in \mathcal{G}} \pi(y) f(y)^2 \right).$$

Using these two lemmas we are now able to bound the second-order moment:

**Proposition 3.10.**

$$\mathbb{E}^{(\pi)} \left( \sum_{i=1}^n f(X_i) \right)^2 \preceq \frac{2\sigma_n^2}{\varepsilon} \cdot I_d, \tag{3.14}$$

*where*

$$\sigma_n^2 = n \cdot \sigma = n \cdot \lambda_{\max} \left( \sum_{y \in \mathcal{G}} \pi(y) f(y)^2 \right).$$

*Proof.* Let us write the sum in Eq. (3.13) as:

$$\mathbb{E}^{(\pi)} \left( \sum_{i=1}^n f(X_i) \right)^2 = \sum_{i=1}^n \sum_{j=1}^i \mathbb{E}^{(\pi)} \left( f(X_i) f(X_j) \right) + \sum_{i=1}^n \sum_{j=i+1}^n \mathbb{E}^{(\pi)} \left( f(X_i) f(X_j) \right)$$

Thanks to lemmas 3.8 and 3.9, we have

$$\begin{aligned} \sum_{j=1}^i \mathbb{E}^{(\pi)} \left( f(X_i) f(X_j) \right) &\preceq \sum_{k=0}^{i-1} \lambda_2^k \sigma^2 \cdot I_d \\ &\preceq \sigma^2 \frac{1 - \lambda_2^{i-1}}{1 - \lambda_2} \cdot I_d \\ &\preceq \sigma^2 \frac{1}{1 - \lambda_2} \cdot I_d \end{aligned}$$

Similarly,

$$\sum_{j=i+1}^n \mathbb{E}^{(\pi)} \left( f(X_i) f(X_j) \right) \preceq \sigma^2 \frac{\lambda_2}{1 - \lambda_2} \cdot I_d$$

Then, by summation, we have that

$$\mathbb{E}^{(\pi)}\left(\sum_{i=1}^{n} f(X_i)\right)^2 \preceq n\sigma^2\left(\frac{1+\lambda_2}{1-\lambda_2}\right)\cdot I_d \preceq \sigma_n^2\cdot\frac{2}{\varepsilon}\cdot I_d$$

□

### 3.3.5 End of the proof

Now, we can use the different lemmas to control the mgf. Starting from Eq. (3.10), we have:

$$\begin{aligned}
\mathbb{P}\left(\lambda_{\max}\left(\sum_{i=1}^{n} f(X_i)\right)\geqslant t\right) &\leqslant \inf_{\theta>0}\left\{e^{-\theta t}\mathbb{E}\operatorname{tr}\exp\left(\theta\sum_{i=1}^{n} f(X_i)\right)\right\}\\
&\leqslant \sup\left(\frac{q_i}{\pi_i}\right)\inf_{\theta>0}\left\{e^{-\theta t}\mathbb{E}^{(\pi)}\operatorname{tr}\exp\left(\theta\sum_{i=1}^{n} f(X_i)\right)\right\} && \text{(lemma 3.7)}\\
&\leqslant \sup\left(\frac{q_i}{\pi_i}\right)\inf_{\theta>0}\left\{e^{-\theta t}\operatorname{tr}\exp\left(g(\theta)\mathbb{E}^{(\pi)}\left(\sum_{i=1}^{n} f(X_i)\right)^2\right)\right\} && \text{(lemma 3.5)}\\
&\leqslant d\cdot\sup\left(\frac{q_i}{\pi_i}\right)\inf_{\theta>0}\left\{\exp\left(-\theta t+g(\theta)\cdot\frac{2\sigma_n^2}{\varepsilon}\right)\right\}. && \text{(proposition 3.10)}
\end{aligned}$$

Following (Tropp, 2012), we notice that the minimal value is reached for $\theta = \log(1 + t\varepsilon/(2\sigma_n^2))$. Let us define $h$ by $h(x) = (1+x)\log(1+x)$, it comes that:

$$\mathbb{P}\left(\lambda_{\max}\left(\sum_{i=1}^{n} f(X_i)\right)\geqslant t\right)\leqslant d\cdot\sup\left(\frac{q_i}{\pi_i}\right)\cdot\exp\left(-\frac{2\sigma_n^2}{\varepsilon}\cdot h\left(\frac{\varepsilon t}{2\sigma_n^2}\right)\right).$$

Proposition 3.1 is obtained with a normalization by $R$. Finally, Corollary 3.2 is a consequence of Bennett's bound since $h(x)\geqslant\frac{x^2/2}{1+x/3}$ for $x\geqslant 0$.

## 3.4 Application to drunk man sampling

In this section, we prove a new Compressed Sensing result, for a sampling scheme based on a random walk on the acquisition space. A CS result was proven in theorem 2.11 page 42. The number of measurements needed to reconstruct any $s$-sparse vector was $\mathcal{O}\left(\frac{s^2}{\varepsilon}K^{*2}(\mathbf{A})\log(2n^2/\eta)\right)$, where $\varepsilon$ is the spectral gap of the chain associated to the

random walk. This bound suffers from the quadratick bottleneck ($s^2$ term). This is a major drawback of this bound, since the number of measurements required to guarantee reconstruction is linear in $s$ in classical CS results in the independent framework. Here we prove that the bound can be improved to $\mathcal{O}\left(\frac{s}{\varepsilon}K^{*2}(\mathbf{A})\log(6n/\eta)\right)$ in a non-uniform framework. In particular, the bound is tight in the independent framework, where $\varepsilon = 1$. The inequality introduced in proposition 3.2 is the cornerstone to prove the following result, using the notation of chapter 2:

**Proposition 3.11.** *Let $x$ be an $s$-sparse random vector such that the signs of its nonzero entries is a Rademacher or Steinhaus sequence. Let $\Omega = X_1, \dots, X_m$ denote a set of $m$ indexes selected using a Markov chain of transition kernel $P$ with gap $\varepsilon$ and $\pi$ its stationary distribution. Assume that $X_1 \sim \pi$. Let $\mathbf{A}_\Omega$ be the matrix obtained by selecting the $m$ lines of $\mathbf{A}$ belonging to the set $\Omega$. Then, if*

$$m \geqslant \frac{14}{3\varepsilon} \cdot K^{*2}(\mathbf{A}) \cdot s \cdot \log^2\left(\frac{6n}{\eta}\right),$$

*with probability $1-\eta$, $x$ is the unique solution of the $\ell_1$ problem (2.1)*

The proof of proposition follows the proof in (Rauhut, 2010). The main difference is that the (Rauhut, 2010, Theorem 7.3) is replaced by:

**Proposition 3.12.** *Let $0 \leqslant \delta \leqslant \frac{1}{2}$. Then, with probability at least*

$$1 - s\exp\left(-\frac{m\delta^2\varepsilon}{\frac{14}{3}K^{*2}(\mathbf{A})s}\right)$$

*the matrix $\mathbf{A}_\Omega$ satisfies*

$$\left\|\frac{1}{m}\mathbf{A}_\Omega^*\mathbf{A}_\Omega - I_n\right\| \leqslant \delta$$

*Proof.* Notice that $\mathbf{A}_\Omega^*\mathbf{A}_\Omega = \sum_{i=1}^m a_{X_i}a_{X_i}^*$. The proposition is a direct consequence of proposition 3.2 with $f : i \in \{1, \dots, n\} \mapsto a_i a_i^* - I_n$. □

A more general CS result could be obtained, without the hypothesis on the sign of the non-zero entries of $x$. Actually, it is possible to improve the result by resorting to the so-called *golfing scheme* technique (Gross, 2011; Candès and Plan, 2011). However, such a proof require other concentration inequalities on sums of vectors with Markovian dependencies, and existing concentration results (Kargin, 2007) should be improved.

## Appendix 1 - Proof of lemma 3.8

In this lemma, we aim at showing that $\mathbb{E}^{(\pi)}[f(X_i)f(X_j)] = \underline{\mathbf{1}}^T F \underline{\Pi P}^{|j-i|} F \underline{\mathbf{1}}$. First, we consider the case $j \geqslant i$. We notice that

$$\begin{aligned}\mathbb{E}^{(\pi)}[f(X_i)f(X_j)] &= \mathbb{E}^{(\pi)}\mathbb{E}^{(\pi)}{}_i[f(X_i)f(X_j)] \\ &= \mathbb{E}^{(\pi)}[f(X_i)\mathbb{E}^{(\pi)}{}_i[f(X_j)]]\end{aligned}$$

Let us show that $\mathbb{E}^{(\pi)}{}_i[f(X_j)] = \underline{e}_{X_i}^T \underline{P}^{j-i} F \underline{\mathbf{1}}$ for $0 \leqslant i \leqslant j$, where $X_i$ is the vertex number reached by the chain after $i$ iterations. For a fixed $j$, let us denote by $(\mathcal{P}_i)$ the assertion:

$$\mathbb{E}^{(\pi)}{}_i[f(X_j)] = \underline{e}_{X_i}^T \underline{P}^{j-i} F \underline{\mathbf{1}}. \qquad (\mathcal{P}_i)$$

Let us show that $(\mathcal{P}_i)$ is true for all $i \leqslant j$. $(\mathcal{P}_j)$ is true, and we assume that $(\mathcal{P}_i)$ is true for a fixed $i$. Let us show that $(\mathcal{P}_{i-1})$ is also true.

$$\begin{aligned}\mathbb{E}^{(\pi)}{}_{i-1}[f(X_j)] &= \sum_{k=1}^{N} P_{X_{i-1},k}\mathbb{E}^{(\pi)}[f(X_j)|X_i = k] \\ &= \sum_{k=1}^{N} e_{X_{i-1}}^T P e_k \underline{e}_k^T \underline{P}^{j-i} F \underline{\mathbf{1}} \\ &= \underline{e}_{X_{i-1}}^T \underline{P}^{j-i+1} F \underline{\mathbf{1}}.\end{aligned}$$

Indeed, $e_{X_{i-1}}^T P e_k \underline{e}_k^T = e_{X_{i-1}}^T P e_k e_k^T \otimes I_d$. Using that $\sum_{k=1}^N e_k e_k^T = I_N$ and the linearity of $\otimes$, it comes that $\sum_{k=1}^N e_{X_{i-1}}^T P e_k \underline{e}_k^T = \underline{e}_{X_{i-1}}^T \underline{P}$, finishing the proof by induction.

Now, notice that $f(X_i) = \underline{e}_{X_i}^T F \underline{e}_{X_i}$. It comes that

$$\begin{aligned}f(X_i)\mathbb{E}^{(\pi)}{}_i[f(X_j)] &= \underline{e}_{X_i}^T F \underline{e}_{X_i} \underline{e}_{X_i}^T \underline{P}^{j-i} F \underline{\mathbf{1}} \\ &= \underline{e}_{X_i}^T F \underline{P}^{j-i} F \underline{\mathbf{1}},\end{aligned}$$

because of the block-diagonal structure of $F$. Finally, $\mathbb{E}^{(\pi)}\underline{e}_{X_i}^T = (\mathbb{E}^{(\pi)} e_{X_i}^T) \otimes I_d = \pi^T \otimes I_d$. Since $\pi^T = \mathbf{1}^T \Pi$ and $I_d = I_d \cdot I_d$, it comes from Prop 3.3 (1) that $\mathbb{E}^{(\pi)}\underline{e}_{X_i}^T = \underline{\mathbf{1}}^T \underline{\Pi}$. The final formula is obtained by noticing that $\underline{\Pi}$ and $F$ commute.

Now, if $j \leqslant i$, we can write:

$$\begin{aligned}
\mathbb{E}^{(\pi)}\left[f(X_i)f(X_j)\right] &= \mathbb{E}^{(\pi)}\left[f(X_i)^* f(X_j)^*\right] \\
&= \mathbb{E}^{(\pi)}\left[(f(X_j)f(X_i))^*\right] \\
&= \left(\mathbb{E}^{(\pi)}\left[f(X_j)f(X_i)\right]\right)^* \qquad \text{Eq. (3.7)}) \\
&= \underline{\mathbf{1}}^T F \underline{\Pi P}^{|j-i|} F \underline{\mathbf{1}},
\end{aligned}$$

since $\underline{P}^*\underline{\Pi} = \underline{\Pi P}$.

## Appendix 2 - Proof of lemma 3.9

First, we notice that $\sigma^2 = \lambda_{\max}\left(\underline{\mathbf{1}}^T F \underline{\Pi} F \underline{\mathbf{1}}\right)$, or again, that $\underline{\mathbf{1}}^T F \underline{\Pi} F \underline{\mathbf{1}} \preceq \sigma^2 \cdot I_d$. Hence, we aim at showing that $\underline{\mathbf{1}}^T F \underline{\Pi P}^k F \underline{\mathbf{1}} \preceq \lambda_2^k \underline{\mathbf{1}}^T F \underline{\Pi} F \underline{\mathbf{1}}$. Now, we can write:

$$\underline{\mathbf{1}}^T F \underline{\Pi P}^k F \underline{\mathbf{1}} - \lambda_2^k \underline{\mathbf{1}}^T F \underline{\Pi} F \underline{\mathbf{1}} = \underline{\mathbf{1}}^T F \underline{\Pi}(\underline{P}^k - \lambda_2^k I_{N\cdot d}) F \underline{\mathbf{1}}$$

Since the chain is reversible, the matrix $\underline{\mathbf{1}}^T F \underline{\Pi}(\underline{P}^k - \lambda_2^k I_{N\cdot d}) F \underline{\mathbf{1}}$ is Hermitian of size $d$. In order to show that it is a negative matrix, we fix $a \in \mathbb{C}^d$ and we show that $a^* \underline{\mathbf{1}}^T F \underline{\Pi}(\underline{P}^k - \lambda_2^k I_{N\cdot d}) F \underline{\mathbf{1}} a \leqslant 0$.

Since $P$ is self-adjoint for $\langle \cdot, \cdot \rangle_\Pi$, there exists $(u_i)_{1 \leqslant i \leqslant N}$ an orthogonal basis of $\mathbb{C}^N$ such that $Pu_i = \lambda_i u_i$, where $\lambda_i$s are sorted in decreasing order as in Eq. (3.8). In particular $(\lambda_1, u_1) = (1, \mathbf{1})$. Hence, $\{u_i \otimes f_j, 1 \leqslant i \leqslant N, 1 \leqslant j \leqslant d\}$ is an orthogonal basis of $\mathbb{C}^{N\cdot d}$ for $\langle \cdot, \cdot \rangle_{\underline{\Pi}}$, and we have:

$$\begin{aligned}
\langle \mathbf{1} \otimes f_i, F \underline{\mathbf{1}} a \rangle_{\underline{\Pi}} &= \sum_{j=1}^{N} (\pi_j f(g_j) a)_i \\
&= 0,
\end{aligned}$$

since $\sum_{j=0}^{N} \pi_j f(g_j) = 0$. Therefore, there is a sequence of complex numbers $(\alpha_{ij})_{2 \leqslant i \leqslant N, 1 \leqslant j \leqslant d}$ such that $F\underline{\mathbf{1}}a = \sum_{i=2}^{N} \sum_{j=1}^{d} \alpha_{ij} u_i \otimes f_j$. We notice that

$$\begin{aligned}
a^* \underline{\mathbf{1}}^T F \underline{\Pi} (\underline{P}^k - \lambda_2^k I_{N \cdot d}) F \underline{\mathbf{1}} a &= \left\langle \sum_{i=2}^{N} \sum_{j=1}^{d} \alpha_{ij} u_i \otimes f_j, (\underline{P}^k - \lambda_2^k I_{N \cdot d}) \sum_{i=2}^{N} \sum_{j=1}^{d} \alpha_{ij} u_i \otimes f_j \right\rangle_{\underline{\Pi}} \\
&= \left\langle \sum_{i=2}^{N} \sum_{j=1}^{d} \alpha_{ij} u_i \otimes f_j, (P^k - \lambda_2^k I_N) \otimes I_d \sum_{i=2}^{N} \sum_{j=1}^{d} \alpha_{ij} u_i \otimes f_j \right\rangle_{\underline{\Pi}} \\
&= \left\langle \sum_{i=2}^{N} \sum_{j=1}^{d} \alpha_{ij} u_i \otimes f_j, \sum_{i=2}^{N} \sum_{j=1}^{d} \alpha_{ij} (\lambda_i^k - \lambda_2^k) u_i \otimes f_j \right\rangle_{\underline{\Pi}} \\
&= \sum_{i=2}^{N} \sum_{j=1}^{d} \alpha_{ij}^2 (\lambda_i^k - \lambda_2^k) \\
&\leqslant 0.
\end{aligned}$$

In other words, $\underline{\mathbf{1}}^T F \underline{\Pi} (\underline{P}^k - \lambda_2^k I_{N \cdot d}) F \underline{\mathbf{1}} \preceq 0$.

# Chapter 4

# A projection algorithm for gradient waveforms design in Magnetic Resonance Imaging

This chapter is based on (Chauffert et al., 2014b).

Collecting the maximal amount of useful information in a given scanning time is a major concern in Magnetic Resonance Imaging (MRI) to speed up image acquisition. The hardware constraints (gradient magnitude, slew rate, ...), physical distortions (e.g., off-resonance effects) and sampling theorems (Shannon, compressed sensing) must be taken into account simultaneously, which makes this problem extremely challenging. To date, the main approach to design gradient waveform has consisted of selecting an initial shape (e.g. spiral, radial lines, ...) and then traversing it as fast as possible. In this paper, we propose an alternative solution: instead of reparameterizing an initial trajectory, we propose to project it onto the convex set of admissible curves. This method has various advantages. First, it better preserves the density of the input curve which is critical in sampling theory. Second, it allows to smooth high curvature areas making the acquisition time shorter in some cases. We develop an efficient iterative algorithm based on convex programming and propose comparisons between the two approaches. For piecewise linear trajectories, our approach generates a gain of scanning time ranging from 20% (echo planar imaging) to 300% (travelling salesman problem) without degrading image quality in terms of signal-to-noise ratio (SNR). For smoother trajectories such as spirals, our method better preserves the sampling density of the input curve, making the sampling pattern relevant for compressed sensing, contrarily to the reparameterization based approaches.

## 4.1 Introduction

The advent of new hardware and sampling theories (e.g., Compressed Sensing or CS) provide unprecedented opportunities to reduce acquisition times in MRI. The design of gradient waveforms minimizing the acquisition time while providing enough information to reconstruct distortion-free images is however an important challenge. Ideally, these two concerns (sampling scheme and gradient waveform design) should be addressed simultaneously, but current theoretical results in sampling theories (either Shannon-based or CS-based) do not permit to incorporate complex physical constraints like the starting position or the traversal speed in $k$-space, despite recent progresses (Chauffert et al., 2014a; Boyer et al., 2015a; Unnikrishnan and Vetterli, 2013; Gröchenig et al., 2014)

To date, the most widespread technique therefore consists of designing gradient waveforms sequentially: a first step aims to find the trajectory support or at least control points, and a second step essentially builds the gradient waveforms to traverse this support or linking these control points. The first step either relies on Shannon sampling theorem (Unnikrishnan and Vetterli, 2013; Gröchenig et al., 2014) or on the concept of variable density sampling (VDS) (Puy et al., 2011; Krahmer and Ward, 2014; Chauffert et al., 2014a). In Shannon theory, the samples located in the $k$-space should lie on a Cartesian grid with a sufficiently small grid step size. A typical instance of such schemes is the echo planar imaging (EPI) trajectory. The wealth of trajectories in VDS is constantly increasing and becomes more and more anchored in theory. It initially started with spirals (Gurney et al., 2006; Pipe and Zwart, 2014) and was progressively enriched with different patterns such as parallel or radial lines (Lustig et al., 2007; Feng et al., 2014), noisy spirals (Lustig et al., 2005), Rosette trajectories (Noll, 1997), shell trajectories (Shu et al., 2006), ... The second step is currently solved by using reparameterization: the goal is to find a feasible waveform traversing the support in the minimum amount of time. This problem can be solved using optimal control (Lustig et al., 2008), convex optimization (Simonetti et al., 1993; Hargreaves et al., 2004), or optimal interpolation of $k$-space control points (Davids et al., 2015). These simple principles however suffer from potentially severe drawbacks. First, reparameterizing the curve changes the density of samples along the curve. This density is now known to be a key aspect in CS (Puy et al., 2011; Chauffert et al., 2014a; Krahmer and Ward, 2014; Adcock et al., 2013), since it directly impacts the number of required measurements to ensure exact recovery (noiseless case) or accurate (noisy case) reconstruction. Second, the challenge of rapid acquisitions is to reduce the scanning time (echo train duration) and limit geometric distortions induced by inhomogeneities of the static magnetic field ($B_0$) by covering the $k$-space as fast as possible. The perfect fit to any arbitrary curve (support constraint) may be time consuming, especially in the high curvature parts of the trajectory. In particular, the

time to traverse piecewise linear trajectories (Chauffert et al., 2014a; Chauffert et al., 2013a; Chauffert et al., 2013c; Wang et al., 2012; Willett., 2011) may become too long. Indeed, the magnetic field gradients have to be set to zero at each singular point of such trajectory. To overcome these two limitations, new gradient waveform design methods have to be pushed forward.

### 4.1.1 Contributions

In this paper, we propose an alternative to reparameterization based on a convex optimization formulation. Given any parameterized curve, our algorithm returns the closest curve that fulfills the gradient constraints. The main advantages of the proposed approach are the following: i) the time to traverse the $k$-space is fixed enabling to find the closest curve in a given time, ii) the distance between the input and output curves is the quantity to be minimized ensuring a low deviation to the original sampling distribution, iii) it is flexible enough to handle additional hardware constraints (e.g., trajectory starting from the $k$-space center, different kinematic constraints,...) in the same framework. We propose an efficient first order dual algorithm to solve the resulting problem and provide theoretical guarantees in terms of convergence rate. We also demonstrate through theory and numerical experiments that the distortion to the initial density is minimized compared to the reparameterization approach. We eventually illustrate the performances of our approach on simulations.

### 4.1.2 Paper organization

In Section 4.2, we review the formulation of MRI acquisition, by recalling the gradient constraints and introducing the projection problem. Then, in Section 4.3, it is shown that curves generated by the proposed strategy (initial parameterization followed by the projection onto the set of physical constraints) may be used to design MRI sampling schemes with locally variable densities. In Section 4.4, we provide an optimization algorithm to solve the projection problem, and estimate its rate of convergence. Next, the behavior of our algorithm is illustrated in Section 4.5 on three complementary cases: one popular sampling scheme, namely EPI trajectory and two VDS strategies (travelling salesman problem or TSP-based curves and spirals), yet advertising the usefulness of the proposed approach for practical MRI applications. The pros and cons of our method are discussed in Section 4.6 and concluding remarks are drawn in Section 4.7.

## 4.2 Design of $k$-space trajectories using physical gradient waveforms.

In this section, we recall the standard modeling of the acquisition constraints in MRI (Hargreaves et al., 2004; Lustig et al., 2008). We justify the lack of accuracy of current reparameterization methods in the VDS context, and motivate the introduction of a new *projection* algorithm that preserves the sampling density.

### 4.2.1 Sampling in MRI

In MRI, images are sampled in the $k$-space domain along parameterized curves $s : [0,T] \mapsto \mathbb{R}^d$ where $d \in \{2,3\}$ denotes the image dimensions. The $i$-th coordinate of $s$ is denoted $s_i$. Let $u : \mathbb{R}^d \to \mathbb{C}$ denote a $d$ dimensional image and $\hat{u}$ be its Fourier transform. Given an image $u$, a curve $s : [0,T] \to \mathbb{R}^d$ and a sampling step $\Delta t$, the image $u$ shall be reconstructed using the set[1]:

$$\mathcal{E}(u,s) = \left\{ \hat{u}(s(j\Delta t)), 0 \leqslant j \leqslant \left\lfloor \frac{T}{\Delta t} \right\rfloor \right\}. \tag{4.1}$$

### 4.2.2 Gradient constraints

The gradient waveform associated with a curve $s$ is defined by $g(t) = \gamma^{-1}\dot{s}(t)$, where $\gamma$ denotes the gyro-magnetic ratio (Hargreaves et al., 2004). The gradient waveforms being obtained by energizing orthogonal gradient coils with electric currents, they are submitted to hardware constraints.

#### 4.2.2.1 Kinematic constraints

Due to physical but also safety (i.e. avoid nerve stimulation) constraints, the electric currents passing through gradient coils have a bounded amplitude and cannot vary too rapidly (slew rate). Mathematically, these constraints read:

$$\|g\| \leqslant G_{\max} \qquad \text{and} \qquad \|\dot{g}\| \leqslant S_{\max}$$

[1] For ease of presentation, we assume that the values of $u$ in the $k$-space correspond to its Fourier transform and we neglect distortions occurring in MRI such as noise. We also neglect the energy decay due to signal relaxation.

where $\|\cdot\|$ denotes either the $\ell^\infty$-norm defined by $\|f\|_\infty := \max\limits_{1\leq i\leq d} \sup\limits_{t\in[0,T]} |f_i(t)|$, or the $\ell^{\infty,2}$-norm defined by $\|f\|_{\infty,2} := \sup\limits_{t\in[0,T]} \left(\sum\limits_{i=1}^{d} |f_i(t)|^2\right)^{\frac{1}{2}}$. These constraints might be Rotation Invariant (RIV) if $\|\cdot\| = \|\cdot\|_{\infty,2}$ or Rotation Variant (RV) if $\|\cdot\| = \|\cdot\|_\infty$, depending on whether each gradient coil is energized independently from others or not. The set of kinematic constraints is denoted $\mathcal{S}$:

$$\mathcal{S} := \left\{ s \in \left(\mathcal{C}^2([0,T])\right)^d, \|\dot{s}\| \leqslant \alpha, \|\ddot{s}\| \leqslant \beta \right\}, \tag{4.2}$$

where $\alpha = \gamma G_{\max}$ and $\beta = \gamma S_{\max}$.

#### 4.2.2.2 Additional affine constraints

Specific MRI acquisitions may require additional constraints, such as:

- Imposing that the trajectory starts from the $k$-space center (i.e., $s(0) = 0$) to save time and avoid blips. The end-point can also be specified by $s(T) = s_T$, if $s_T$ can be reached during travel time $T$.
- In the context of multi-shot MRI acquisition, several radio-frequency pulses are necessary to cover the whole $k$-space. Hence, it makes sense to enforce the trajectory to start from the $k$-space center at every $TR$ (repetition time)[2]: $s(m \cdot TR) = 0, 0 \leqslant m \leqslant \left\lfloor \frac{T}{TR} \right\rfloor$.
- In addition to starting from the $k$-space center, one could impose the initial speed as for instance: $\dot{s}(0) = 0$.
- To avoid artifacts due to flow motion in the object of interest, gradient moment nulling (GMN) techniques have been introduced in (Majewski et al., 2010) for spin or gradient echo sequences. In terms of constraints, nulling the $i^{\text{th}}$ moment reads $\int_{t=0}^{TE} t^i g(t) \mathrm{d}t = 0$, where $TE$ denotes the echo time. For example, cancelling the first-order moment compensates the motion of spins moving with constant speed.

Each of these constraints can be modelled by an affine relationship. Hereafter, the set of affine constraints is denoted by $\mathcal{A}$:

$$\mathcal{A} := \left\{ s : [0,T] \to \mathbb{R}^d, A(s) = v \right\},$$

where $v$ is a vector of parameters in $\mathbb{R}^p$ ($p$ is the number of additional constraints) and $A$ is a linear mapping from the curves space to $\mathbb{R}^p$.

[2] corresponding to the delivery of every radio-frequency pulse.

A sampling trajectory $s : [0, T] \to \mathbb{R}^d$ will be said to be *admissible* if it belongs to the set $\mathcal{S} \cap \mathcal{A}$. In what follows, we assume that this set is non-empty, i.e. $\mathcal{S} \cap \mathcal{A} \neq \emptyset$. Moreover, we assume, without loss of generality, that the linear constraints are independent (otherwise some could be removed).

### 4.2.3 Finding an optimal reparameterization

The traditional approach to design an admissible curve $s \in \mathcal{S}$ given an arbitrary curve $c : [0, T] \to \mathbb{R}^d$ consists of finding a reparameterization $p$ such that $s = c \circ p$ satisfies the physical constraints while minimizing the acquisition time. This problem can be cast as follows:

$$T_{\texttt{Rep}} = \min T' \text{ such that } \exists\, p : [0, T'] \mapsto [0, T],\ c \circ p \in \mathcal{S}. \tag{4.3}$$

It can be solved efficiently using optimal control (Lustig et al., 2008) or convex optimization (Hargreaves et al., 2004). The resulting solution $s = c \circ p$ has the same support as $c$. This method however suffers from an important drawback when used in the CS framework: it does not provide any control on the density of samples along the curve. For example, for a given curve support shown in Fig. 4.1(a), we illustrate the new parameterization (keeping the same support) and the corresponding magnetic field gradients (see Fig. 4.1(b) for a discretization of the curve and (c) for the gradient profile). We notice that the new parameterized curve has to stop at every angular point of the trajectory, yielding more time spent by the curve in the neighbourhood of these points (and more points in the discretization of the curve in Fig. 4.1(b)). This phenomenon is likely to modify the sampling distribution, as illustrated in Section 4.3.

The next part is dedicated to introducing an alternative method relaxing the constraint of keeping the same support as $c$.

### 4.2.4 Projection onto the set of constraints

Here, we propose to find the projection of the given input curve $c$ onto the set of admissible curves $\mathcal{S}$:

$$s^* := \underset{s \in \mathcal{S} \cap \mathcal{A}}{\operatorname{argmin}} \frac{1}{2} d^2(s, c) = \underset{s \in \mathcal{S} \cap \mathcal{A}}{\operatorname{argmin}} \frac{1}{2} \|s - c\|_2^2 \tag{4.4}$$

where $d^2(s, c) = \|s - c\|_2^2 := \int_{t=0}^{T} \|s(t) - c(t)\|_2^2 \, dt$. This method presents important differences compared to the above mentioned optimal control approach: i) the solution $s^*$ and $c$ have different support (see Fig. 4.1(d)) unless $c$ is admissible; ii) the sets

composed of the discretization of $c$ and $s^*$ at a given sampling rate are close to each other (Fig. 4.1(e)); iii) the acquisition time $T$ is fixed and equal to that of the input curve $c$. Time to traverse a curve is generally different from optimal reparameterization. In particular for piecewise linear curves, it is generally lower (see Fig. 4.1(f) where $T < T_{\texttt{Rep}}$).

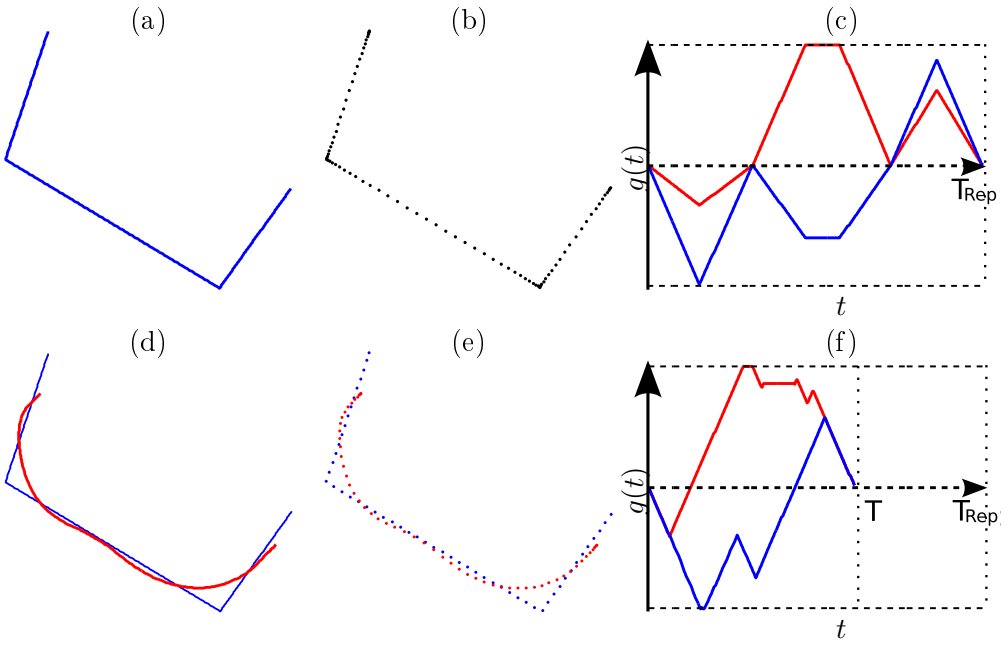


FIGURE 4.1: Comparison of two methods to design gradient waveforms. **Top row:** Optimal control-based parameterization (Lustig et al., 2008). (a): input curve support. (b): discrete representation of the optimal reparameterization of the curve in $\mathcal{S}$. (c): corresponding gradient waveforms $(g_x, g_y)$. Dashed lines correspond to 0 and +/- $G_{\max}$. **Bottom row:** Illustration of the projection algorithm. (d): same input curve $c$ as in (a) parameterized at maximal speed, and the support of the projected curve $s^*$ onto $\mathcal{S}$. (e): discrete representation of the input and projected curves. (f): corresponding gradient waveforms $(g_x, g_y)$ with the same time scale as in (c): the time to traverse the $s^*$ is *39%* shorter.

In the next section, we explain why the empirical distribution of the samples along the projected curve is closer to that of points lying on the input curve. Also, we illustrate how the parameterization can distort the sampling distribution.

## 4.3 Control of the sampling density

Recent works have emphasized the importance of the sampling density (Chauffert et al., 2014a; Puy et al., 2011; Krahmer and Ward, 2014; Adcock et al., 2013) in the CS-MRI framework, i.e. in an attempt to reduce the amount of acquired data while preserving image quality at the reconstruction step. The choice of an accurate sampling distribution is crucial since it directly impacts the number of required measurements (Rauhut, 2010). In this paper, we will denote by $\pi$ a distribution defined over the $k$-space $K$. The profile

of this distribution can be obtained by theoretical arguments (Chauffert et al., 2014a; Puy et al., 2011; Krahmer and Ward, 2014; Adcock et al., 2013) leading to distributions as the one depicted in Fig. 4.2(a). Some heuristic distributions (e.g., radial) are known to perform well in CS-MRI experiments (Fig. 4.2(b)). A comparison between these two approaches can be found in (Chauffert et al., 2013b).

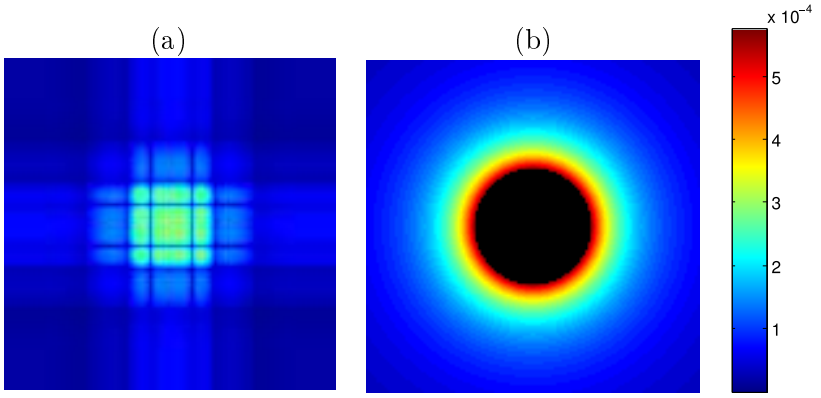


FIGURE 4.2: Examples of 2D sampling distribution. (a): optimal distribution for a Symmlet transform (Puy et al., 2011; Chauffert et al., 2014a). (b): radial distribution advocated in (Krahmer and Ward, 2014; Chauffert et al., 2014a): $p(k) \propto 1/|k|^2$.

However, designing a trajectory that performs sampling according to a fixed distribution while satisfying gradient constraints is really challenging and has not been addressed so far. The classical approach consists of:

1. Finding an input curve (admissible or not) $c$ with good distribution; We provide various strategies to achieve this step in Appendix 4.7.

2. Estimating the fastest reparameterization of $c$ that belongs to the set of constraints.

In this paper, we suggest to replace the second step by:

2') Estimating $s^*$ the projection of $c$ onto the set of constraints, by solving Eq. (4.4).

We show that step 2') is preferable to step 2) since it better preserves the sampling density (or empirical measure). We begin by showing it through a theoretical study in paragraph 4.3.1 and then validate it through numerical experiments in paragraph 4.3.2. The reader not interested by theoretical arguments can go directly to Subsection 4.3.2.

### 4.3.1 Theoretical study of the density control

To formalize the notion of density, we need to introduce the definition of the empirical distribution of a curve.

**Definition 4.1** (Empirical measure of a curve)**.** Let $\lambda$ denote the Lebesgue measure and $\lambda_T = \frac{\lambda}{T}$ denote the Lebesgue measure normalized on the interval $[0, T]$. The *empirical measure* of a curve $s : [0, T] \mapsto K \subseteq \mathbb{R}^d$ is defined for any measurable set $\omega$ of $K$ as:

$$P_s(\omega) = \lambda_T(s^{-1}(\omega)).$$

This definition means that the mass of a set $\omega$ is proportional to the time spent by the curve in $\omega$.

To measure the distortion between an input curve and the projected one, we need to design a distance between measures. In this work, we propose to use the Wasserstein distance $W_2$ defined hereafter:

**Definition 4.2** (Wasserstein distance $W_2$)**.** Let $M$ be a domain of $\mathbb{R}^d$ and $\mathcal{P}(M)$ be the set of measures over $M$. For $\mu, \nu \in \mathcal{P}(M)$, $W_2$ is defined as:

$$W_2(\mu, \nu) = \left( \inf_{\sigma \in \Pi(\mu,\nu)} \int \|x - y\|_2^2 \mathrm{d}\sigma(x, y) \right)^{\frac{1}{2}} \tag{4.5}$$

where $\Pi \subset \mathcal{P}(M \times M)$ denote the set of measures over $M \times M$ with marginals $\mu$ and $\nu$ on the first and second factors, respectively.

$W_2$ is a distance over $\mathcal{P}(M)$ (see e.g., (Villani, 2008)). Intuitively, if $\mu$ and $\nu$ are seen as mountains, the distance is the minimum cost of moving the mountains of $\mu$ into the mountains of $\nu$, where the cost is the $\ell_2$-distance of transportation multiplied by the mass moved. Hence, the coupling $\sigma$ encodes the deformation map to turn one distribution ($\mu$) into the other ($\nu$).

Let us now analyze the distortion between the empirical distribution of the projected curve $P_{s^*}$ and the target distribution $\pi$. Since $W_2$ is a distance between measures, the triangle inequality holds:

$$W_2(P_{s^*}, \pi) \leqslant \underbrace{W_2(P_c, \pi)}_{\text{Initial distortion}} + \underbrace{W_2(P_{s^*}, P_c)}_{\text{Projection distortion}} . \tag{4.6}$$

The deviation is controlled by two terms: the *initial distortion* term $W_2(P_c, \pi)$ and the *projection distortion* term $W_2(P_{s^*}, P_c)$. The first term depends of the choice of the input curve $c$. This choice is crucial but is out the scope of this paper since it is not directly related to gradient waveform design. We still show in Appendix A that this term can be controlled precisely in a few cases of interest (spiral, TSP).

We are now interested in controlling the *Projection distortion* term $W_2(P_{s^*}, P_c)$. The following proposition shows that the $W_2$ distance between the empirical distributions of

the input and output curves ($c$ and $s^*$, respectively) is controlled by the quantity $d(s^*, c)$ to be minimized when solving Eq. (4.4).

**Proposition 4.3.** *For any two curves $s$ and $c : [0, T] \to \mathbb{R}^d$:*

$$W_2(P_s, P_c) \leqslant d(s, c).$$

*Proof.* In terms of distributions, the quantity $d(s, c)$ reads:

$$d^2(s, c) = \int_{M \times M} \|x - y\|_2^2 \mathrm{d}\sigma_{s,c}(x, y) \tag{4.7}$$

where $\sigma_{s,c}$ is the coupling between the empirical measures $P_s$ and $P_c$ defined for all couples of measure sets $(\omega_1, \omega_2) \in M^2$ by $\sigma_{s,c}(\omega_1, \omega_2) = \frac{1}{T} \int_{t=0}^{T} 1_{\omega_1}(s(t)) 1_{\omega_2}(c(t)) \mathrm{d}t$, where $1_\omega$ denote the indicator function of $\omega$. The choice of this coupling is equivalent to choosing the transformation map as the association of locations of $c(t)$ and $s(t)$ for every $t$. We notice that the quantity to be minimized in Eq. (4.7) is an upper bound of $W_2(P_s, P_c)^2$, with the specific coupling $\sigma_{s,c}$. □

To sum up, solving the projection problem (4.4) and finding $s^*$ amounts to minimizing an upper-bound of $W_2(P_{s^*}, \pi)$, the Wasserstein distance between the target density $\pi$ and the empirical distribution $P_{s^*}$, if we neglect the influence of the *initial parameterization* $c$. In some sense, our projection algorithm is therefore the best way to obtain a feasible curve and to preserve the input curve empirical measure. As will be seen in the next paragraph, densities are indeed much better preserved using projections than reparameterizations.

### 4.3.2 Numerical study of the density control

Next, we performed simulations to show that the sampling density is better preserved using our algorithm compared to the optimal control approach. For doing so, we use travelling salesman-based (TSP) sampling trajectories (Chauffert et al., 2013c; Chauffert et al., 2014a), which are an original way to design random trajectories which empirical distribution is any target density $\pi$ such as the one represented in Fig. 4.2(a). $10,000$ such independent TSP were drawn and parameterized with arc-length: note that these parameterizations are not admissible in general. Then, we sampled each trajectory at constant rate $\Delta t$ (as in Fig. 4.3 (top-row, left)), to form an histogram depicting the empirical distribution shown in Fig. 4.3 (top-row, center). The latter was eventually compared to $\pi$ in Fig. 4.3 (top-row, right). It is worth noting that the error was actually not close to zero, since the convergence result enounced in (Chauffert et al., 2014a) is

asymptotic, i.e. when the length of the TSP curve tends to infinity whereas the latter remains bounded in this experiment.

In Fig. 4.3 (second row), we show that the classical reparameterization technique (Lustig et al., 2008) leads to a major distortion of the sampling density, because of its behavior on the angular points already illustrated in Fig. 4.1(b). Then, we considered three constant speed parameterizations and projected them onto the same set of constraints ($G_{\max} = 40$ mT.m$^{-1}$ and $S_{\max} = 150$ mT.m$^{-1}$.ms$^{-1}$). Among these three initial candidates, we started by using an initial parameterization with low velocity (10 % of the maximal speed $\gamma G_{\max}$ with $\gamma = 42.576$ MHz.T$^{-1}$ for proton imaging), which projection fits the sampling density quite well. Then, we increased the velocity to progressively reach 50 % and even 100 % of the maximal speed. The distortion of the sampling density of the projected curve increased, but remained negligible in contrast to what we observed for the exact reparameterization. Hence, this example illustrates that starting from a continuous trajectory whose an empirical sampling distribution is close to the target $\pi$, our projection algorithm yields feasible gradient waveforms while sampling the $k$-space along a discretized trajectory whose empirical density is close to $\pi$ too.

## 4.4 Finding feasible waveforms using convex optimization

Since the set of constraints $\mathcal{S} \cap \mathcal{A}$ is convex, closed and non-empty, Problem (4.4) always admits a unique solution. Even though $\mathcal{S}$ has a rather simple structure[3], it is unlikely that an explicit solution to Problem (4.4) can be found. In what follows, we thus propose a numerical algorithm to compute the projection.

### Problem discretization

A discrete-time curve $\boldsymbol{s}$ is defined as a vector in $\mathbb{R}^{n \cdot d}$ where $n$ is the number of time points. Let $\boldsymbol{s}(i) \in \mathbb{R}^d$ denote the curve location at time $(i-1)\delta t$ with $\delta t = \frac{T}{n-1}$. The discrete-time derivative $\dot{\boldsymbol{s}} \in \mathbb{R}^{n \cdot d}$ is defined using first-order differences:

$$\dot{\boldsymbol{s}}(i) = \begin{cases} 0 & \text{if } i = 1, \\ (\boldsymbol{s}(i) - \boldsymbol{s}(i-1))/\delta t & \text{if } i \in \{2, \dots, n\}. \end{cases}$$

In the discrete setting, the first-order differential operator can be represented by a matrix $\dot{\mathbf{M}} \in \mathbb{R}^{n \cdot d \times n \cdot d}$, i.e. $\dot{\boldsymbol{s}} = \dot{\mathbf{M}}\boldsymbol{s}$. We define the discrete second-order differential operator by $\ddot{\mathbf{M}} = -\dot{\mathbf{M}}^* \dot{\mathbf{M}} \in \mathbb{R}^{n \cdot d \times n \cdot d}$.

[3] it is just a polytope when the $\ell^\infty$-norm is used.

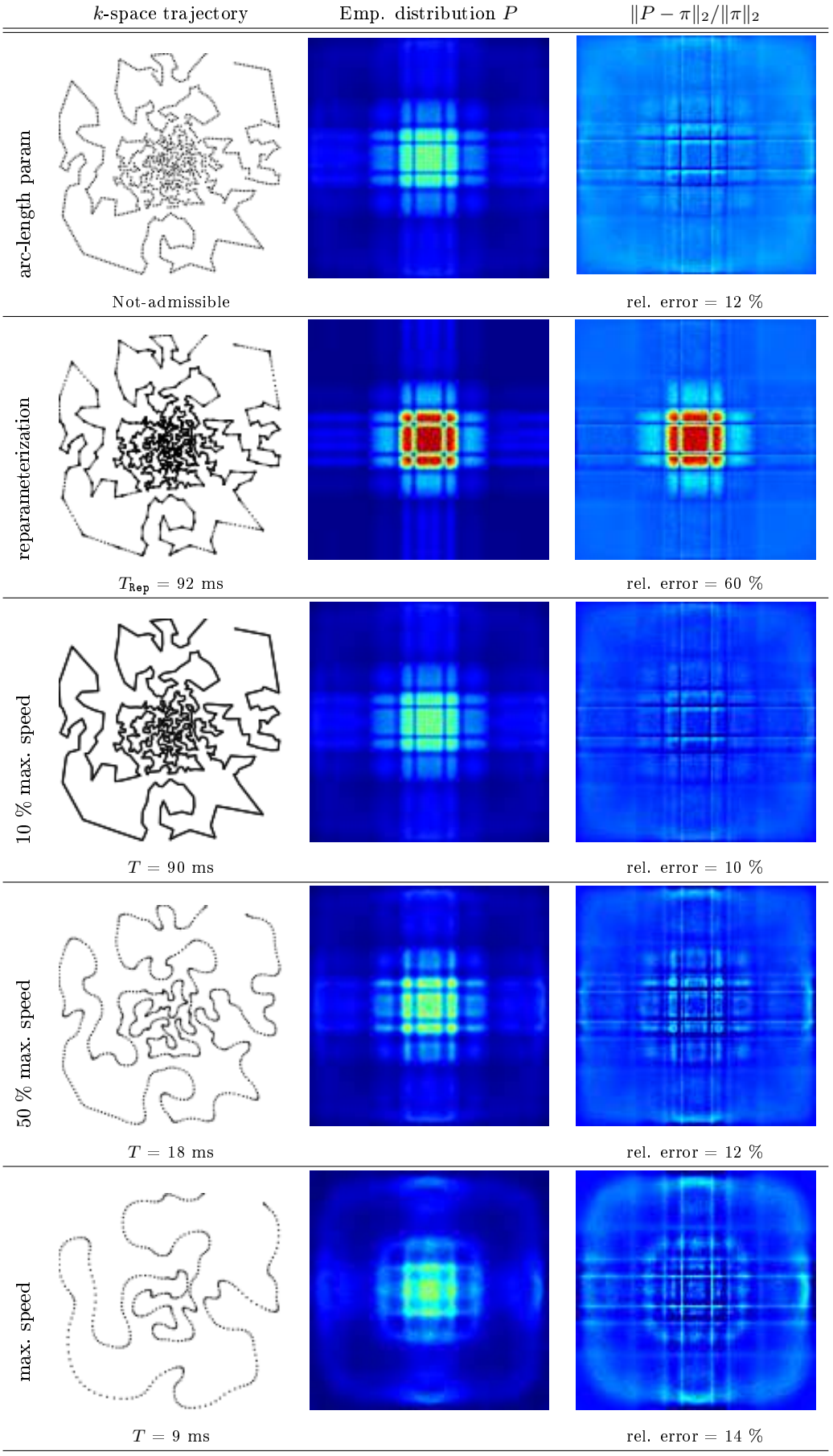


FIGURE 4.3: Illustration of TSP trajectories traversed with arc-length parameterization (**top row**), optimal control (**second row**) and with our projection algorithm (**rows 3-5**). Columns represent the $k$-space trajectory (**left**), the empirical distribution $P$ (**center**) and the difference with the target distribution $\pi$ shown in Fig. 4.2(a) (**right**). At the bottom, the relative error $\|P-\pi\|_2/\|\pi\|_2$ between the two is reported.

### An efficient projection algorithm

The discrete primal problem we consider is the same as (4.4) except that all objects are discretized. It reads:

$$\min_{\boldsymbol{s}\in\boldsymbol{\mathcal{S}}\cap\boldsymbol{\mathcal{A}}} \frac{1}{2}\|\boldsymbol{s}-\boldsymbol{c}\|_2^2, \tag{$\mathcal{P}$}$$

where $\boldsymbol{\mathcal{S}} := \{\boldsymbol{s}\in\mathbb{R}^{n\cdot d}, \|\dot{\mathbf{M}}\boldsymbol{s}\|\leqslant\alpha, \|\ddot{\mathbf{M}}\boldsymbol{s}\|\leqslant\beta\}$ with all norms discretized, and $\boldsymbol{\mathcal{A}}$ are the discrete counterparts of $\mathcal{S}$ and $\mathcal{A}$, respectively. Next, the main idea is to take advantage of the structure of the dual problem of $\mathcal{P}$ to design an efficient projection algorithm. The following proposition specifies this dual problem and the primal-dual relationships.

**Proposition 4.4.** *Let* $\|\boldsymbol{q}\|_* := \sup_{\|\boldsymbol{s}\|\leq 1}\langle\boldsymbol{s},\boldsymbol{q}\rangle$ *denote the dual norm of* $\|\cdot\|$. *The following equality holds:*

$$\min_{\boldsymbol{s}\in\boldsymbol{\mathcal{S}}\cap\boldsymbol{\mathcal{A}}} \frac{1}{2}\|\boldsymbol{s}-\boldsymbol{c}\|_2^2 = \sup_{\boldsymbol{q}_1,\boldsymbol{q}_2\in\mathbb{R}^{n\cdot d}} F(\boldsymbol{q}_1,\boldsymbol{q}_2) - \alpha\|\boldsymbol{q}_1\|_* - \beta\|\boldsymbol{q}_2\|_*, \tag{4.8}$$

*where*

$$F(\boldsymbol{q}_1,\boldsymbol{q}_2) = \min_{\boldsymbol{s}\in\boldsymbol{\mathcal{A}}} \langle\dot{\mathbf{M}}\boldsymbol{s},q_1\rangle + \langle\ddot{\mathbf{M}}\boldsymbol{s},q_2\rangle + \frac{1}{2}\|\boldsymbol{s}-\boldsymbol{c}\|_2^2. \tag{4.9}$$

*Moreover, let* $(\boldsymbol{q}_1^*,\boldsymbol{q}_2^*)$ *denote any minimizer of the dual problem* (4.8), $\boldsymbol{s}^*$ *denote the unique solution of the primal problem* ($\mathcal{P}$) *and* $\boldsymbol{s}^*(\boldsymbol{q}_1^*,\boldsymbol{q}_2^*)$ *denote the solution of the minimization problem* (4.9). *Then* $\boldsymbol{s}^* = \boldsymbol{s}^*(\boldsymbol{q}_1^*,\boldsymbol{q}_2^*)$.

*Proof.* The proof is given in Appendix B. □

The following proposition gives an explicit expression of $\boldsymbol{s}^*(\boldsymbol{q}_1^*,\boldsymbol{q}_2^*)$.

**Proposition 4.5.** *The minimizer*

$$\boldsymbol{s}^*(\boldsymbol{q}_1^*,\boldsymbol{q}_2^*) = \arg\min_{\boldsymbol{s}\in\boldsymbol{\mathcal{A}}} \langle\dot{\mathbf{M}}\boldsymbol{s},\boldsymbol{q}_1\rangle + \langle\ddot{\mathbf{M}}\boldsymbol{s},\boldsymbol{q}_2\rangle + \frac{1}{2}\|\boldsymbol{s}-\boldsymbol{c}\|_2^2$$

*is given by*

$$\boldsymbol{s}^*(\boldsymbol{q}_1,\boldsymbol{q}_2) = \boldsymbol{z} + \boldsymbol{A}^+(\boldsymbol{v}-\boldsymbol{A}\boldsymbol{z}), \tag{4.10}$$

*where* $\boldsymbol{A}\in\mathbb{R}^{p\times n\cdot d}$ *is a matrix encoding the affine constraints, and* $\boldsymbol{A}^+ = \boldsymbol{A}^*(\boldsymbol{A}\boldsymbol{A}^*)^{-1}$ *denotes its pseudo-inverse*[4]. *In addition,* $\boldsymbol{z} = \boldsymbol{c} - \dot{\mathbf{M}}^*\boldsymbol{q}_1 - \ddot{\mathbf{M}}^*\boldsymbol{q}_2$.

*Proof.* The proof is given in Appendix C. □

[4]Since the constraints are supposed to be linearly independent, $\boldsymbol{A}^+$ is well-defined.

Let us now analyse the smoothness properties of $F$.

**Proposition 4.6.** *Function $F(\boldsymbol{q}_1, \boldsymbol{q}_2)$ is concave differentiable with gradient given by*

$$\nabla F(\boldsymbol{q}_1, \boldsymbol{q}_2) = -\begin{pmatrix} \dot{\mathbf{M}}\boldsymbol{s}^*(\boldsymbol{q}_1, \boldsymbol{q}_2) \\ \ddot{\mathbf{M}}\boldsymbol{s}^*(\boldsymbol{q}_1, \boldsymbol{q}_2) \end{pmatrix}. \tag{4.11}$$

*Moreover, the gradient mapping $\nabla F$ is Lipschitz continuous with constant $L = |||\dot{\mathbf{M}}^*\dot{\mathbf{M}} + \ddot{\mathbf{M}}^*\ddot{\mathbf{M}}|||$, where $|||\mathbf{M}|||$ denotes the spectral norm of $\mathbf{M}$.*

Proposition 4.6 is a direct application of (Nesterov, 2005, Theorem 1) (see also (Hiriart-Urruty and Lemaréchal, 1996)). The dual problem (4.8) has a favorable structure for its optimization: it is the sum of a differentiable convex function $\tilde{F}(\boldsymbol{q}_1, \boldsymbol{q}_2) = -F(\boldsymbol{q}_1, \boldsymbol{q}_2)$ and of a simple convex function $G(\boldsymbol{q}_1, \boldsymbol{q}_2) = \alpha\|\boldsymbol{q}_1\|_* + \beta\|\boldsymbol{q}_2\|_*$. The sum $\tilde{F} + G$ can thus be minimized efficiently using accelerated proximal gradient descents (Nesterov, 1983) (see Algorithm 1 below).

**Algorithm 1:** Projection algorithm in the dual space

**Input**: $\boldsymbol{c} \in \mathbb{R}^{n\cdot d}$, $\alpha, \beta > 0$, $n_{it}$.
**Output**: $\tilde{\boldsymbol{s}} \in \mathbb{R}^{n\cdot d}$ an approximation of the solution $\boldsymbol{s}^*$.
**Initialize** $\boldsymbol{q}^{(0)} = (\boldsymbol{q}_1^{(0)}, \boldsymbol{q}_2^{(0)})$ with $\boldsymbol{q}_i^{(0)} = 0$ for $i = 1, 2$. Set $\boldsymbol{y}^{(0)} = \boldsymbol{q}^{(0)}$.
Set $\ell = 1/L$.
**for** $k = 1 \dots n_{it}$ **do**
  $\boldsymbol{q}^{(k)} = \mathbf{prox}_{\ell G}(\boldsymbol{y}^{(k-1)} - \ell\nabla\tilde{F}(\boldsymbol{y}^{(k-1)}))$
  $\boldsymbol{y}^{(k)} = \boldsymbol{q}^{(k)} + \frac{k-1}{k+2}(\boldsymbol{q}^{(k)} - \boldsymbol{q}^{(k-1)})$
**return** $\tilde{\boldsymbol{s}} = \boldsymbol{s}^*\left(\boldsymbol{q}_1^{(n_{it})}, \boldsymbol{q}_2^{(n_{it})}\right)$.

Moreover, by combining the convergence rate results of (Nesterov, 1983; Beck and Teboulle, 2009b) and some convex analysis (see Appendix D), we obtain the following convergence rate:

**Theorem 4.7.** *Algorithm 1 ensures that the distance to the minimizer decreases as $\mathcal{O}\left(\frac{1}{k^2}\right)$:*

$$\|\boldsymbol{s}^{(k)} - \boldsymbol{s}^*\|_2^2 \leq \frac{2L\|\boldsymbol{q}^{(0)} - \boldsymbol{q}^*\|_2^2}{k^2}. \tag{4.12}$$

## 4.5 Numerical experiments

To compare our results with (Lustig et al., 2008), we used the same gradient constraints. In particular, the maximal gradient norm $G_{\max}$ was set to 40 $\mathrm{mT.m^{-1}}$, and

the slew-rate $S_{\max}$ to 150 $\text{mT.m}^{-1}\text{.ms}^{-1}$. We assume that the constraints are Rotation Invariant (RIV). The image field of view (FOV) is assumed to be 20 cm and $K_{\max} = N/(2 \cdot FOV)$ where $N$ is the target spatial grid size for image reconstruction. The sampling rate was fixed to $\Delta t = 4$ $\mu$s except for spiral imaging. For the ease of trajectory representation, we limit ourselves to 2D sampling curves, although our algorithm encompasses the 3D setting.

The Matlab codes embedding the projection algorithm as well as the scripts to reproduce the results depicted hereafter are available at `http://chauffertn.free.fr/codes.html`. Hereafter, the supplementary affine constraints (e.g., nulling moments) are not taken into account. However, they have been implemented in the code so that every end-user can play with. Simulations were performed on a Linux Ubuntu (64 bits) workstation with an Intel Xeon(R) CPU E5-2630 v2 @2.60GHz processor and 64 GB of RAM. The computation time required to run the experiments range from 2 min. (EPI with $17,225$ points) to 4 min. (TSP trajectory with $45,000$ points) and the number of iterations of Algorithm 1 to achieve convergence was $15,000$, to satisfy $||(s^{(k+1)} - s^{(k)})/s^{(k)}|| < 10^{-3}$.

To measure the impact of the proposed projection algorithm and compare it with the optimal reparameterization, we also performed image reconstruction and computed image quality in terms of Signal-to-Noise-Ratio (SNR). To this end, we performed simulations by starting from a high-resolution $N \times N$ MRI phantom ($N = 1024$) depicted in Fig. 4.4. Next, we massively undersampled its Fourier transform by the two competing sampling strategies and analyzed image quality after non-Cartesian reconstruction. For the sake of self-containedness, all investigated trajectories are depicted in Fig. 4.7 and quantitative results corresponding traversal times and SNR of reconstructed images are reported in Tab. 4.1. In what follows, we first discuss the results of our method in the context

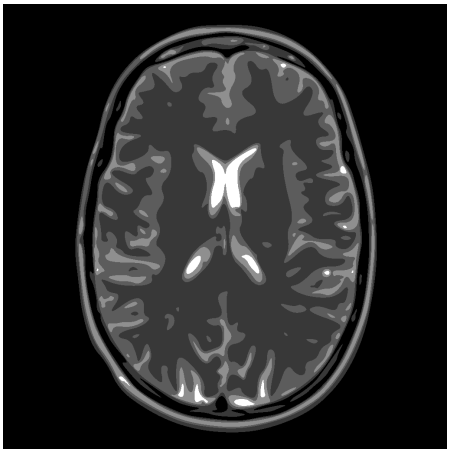

FIGURE 4.4: MRI phantom of size $N \times N$ ($N = 1024$) used for the experiments.

of classical (piecewise linear) EPI trajectory. Then, we illustrate the behavior of our

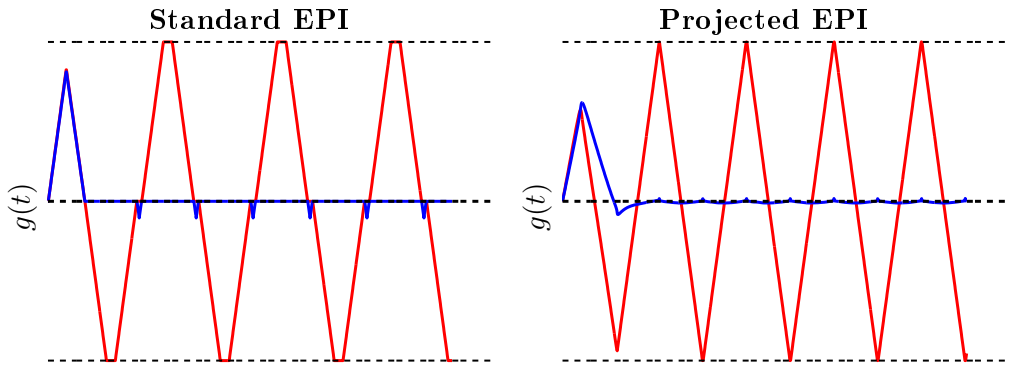


FIGURE 4.5: Comparison between magnetic field gradients $g = (g_x, g_y)$ during *the first 5 ms* for standard EPI trajectory (left) and projected EPI (right).

algorithm on two VDS: classical spiral (smooth) trajectories and TSP-based (piecewise linear) trajectories.

### 4.5.1 EPI trajectories

EPI trajectories are a classical way of probing the $k$-space on a 2D-Cartesian grid. We compared a standard EPI with ramp-sampling (a sample was measured every $\Delta t$ from $t = 0$ to $T_{\texttt{Rep}}$) on $N = 128$ lines, parameterized with optimal control and a trajectory that traverses the $k$-space at constant speed (70% of the maximal gradient intensity), projected onto $\boldsymbol{\mathcal{S}}$ using our algorithm.

As shown in Fig. 4.7 (third row), the projected trajectory has a smaller support than standard EPI. In particular, the resolution in the readout direction is slightly decreased. However, the time to traverse $k$-space is shorter ($T = 68.9$ ms) using our algorithm as compared to the EPI trajectory ($T_{\texttt{Rep}} = 89.6$ ms). To provide a better insight on this acceleration factor, we depict in Fig. 4.5 the first 5 ms of the gradient waveforms for the two approaches. The corresponding acquired lines are colored in red in Fig. 4.7. While standard EPI is able to acquire 6.5 lines (Fig. 4.5-left) in this amount of time, the projected trajectory achieves the extended coverage of 8.5 lines (Fig. 4.5-right). Indeed, gradient blips are smoothed providing a substantial time reduction. In terms of image quality, we observed that the degradation of resolution along one direction has no significant impact since the SNR of reconstructed image is higher for the projected trajectory compared to standard EPI (Tab. 4.1). Hence, in the EPI context, the projection algorithm allows us to traverse the $k$-space faster without degrading the image quality.

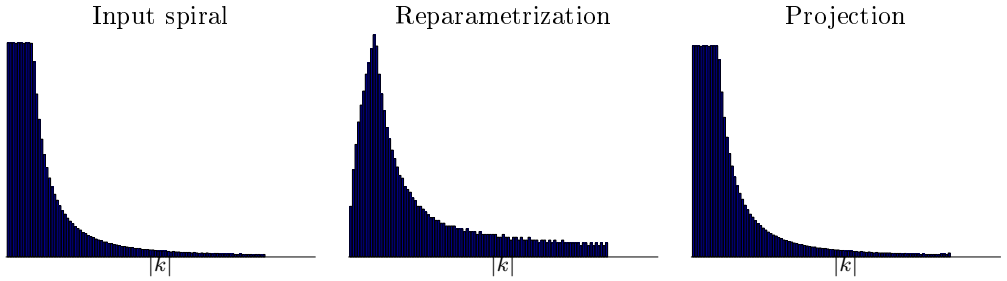


FIGURE 4.6: Decay of the spiral for an input spiral with density $\pi(k) \propto 1/|k|^2$. Histogram of the values $r(t)$ for input spiral (not admissible), optimal reparametrization, and projection.

### 4.5.2 Spiral trajectories

The case of spiral trajectories is more tricky as explained below. For any radial density $\pi$, there exists a spiral that performs $k$-space sampling according to $\pi$. This trajectory is parameterized by $c(t) = r(t/T)\exp(i2\pi n\theta(t/T))$ and thus controlled by its time-varying modulus $r(t)$ and phase $\theta(t)$ and by the number of revolutions $n \in \mathbb{R}_+$ over the fixed traversal time $T$. The relation between $\pi$ and $r(\cdot)$ is given in Appendix A, Eq. (4.14), hence the choice of $r(\cdot)$ determines $\pi$, whereas $\theta(\cdot)$ and $n$ control the shape of the spiral. For fixed $T$ and $r(\cdot)$, finding $\theta(\cdot)$ and $n$ such that the spiral is optimal in the sense that the kinematics constraints $\boldsymbol{\mathcal{S}}$ are saturated, is an open issue. Indeed, in the literature (Kim et al., 2003), it has been shown that different types of gradient parameterizations may yield different sampling patterns, hence various $\pi$. However, to the best of our knowledge, the inverse problem which consists of inferring the parameterization from the target density $\pi$, has never been solved.

Here, we provide a partial solution that relies on two ingredients: first, setting the function $r(\cdot)$ according to (4.14) and second choosing a constant angular speed $\omega$ such that $\theta(t) = \omega t$. This approach actually remains suboptimal since considering a constant $\omega$ imposes too low gradient magnitudes at the beginning of the trajectory (i.e. for the $k$-space center). The pair $(\omega, n)$ must satisfy the constraints in $\boldsymbol{\mathcal{S}}$. For instance, to saturate the magnitude gradient constraint one may choose $(\omega, n)$ such that: $2\pi n\omega K_{\max} = \gamma G_{\max}$.

In our experiment, we adopted this strategy for the above defined $G_{\max}$ and the selected $K_{\max}$ (see Fig. 4.7). We also set $T = 200$ ms and $\Delta t = 24$ $\mu$s, in order to meet an additional memory size constraint[5]. In Fig. 4.6, we illustrate how the sampling density $\pi(k) \propto 1/|k|^2$ is impacted after optimal control reparametrization whereas it is preserved when applying our projection algorithm. The histogram peak associated with the reparameterization is shifted to the right i.e. towards high frequencies meaning that the low frequencies are undersampled. This is the direct consequence of using a too fast traversal

[5]the buffer size of the analog-to-digital converter is 8912 in standard MRI scanners.

speed (see Fig. 4.7: the samples of the spiral are more spaced in the reparameterization scenario). The traversal time of the spiral is indeed $T_{\texttt{Rep}} = 42$ ms with reparametrization and $T$ after projection (Tab. 4.1). This also explains the significant difference of image quality by almost 5 dB in favor of the projection approach.
On the contrary, if the initial parameterization is not admissible ($\omega$ too large), we observed that the output trajectory of the projection algorithm concentrates on concentric circles corresponding to the maximal speed allowed by the gradient magnitude constraint (results not shown). In contrast, the optimal reparameterization is not impacted since it only depends on the support of the spiral. Hence, the choice of the initial parametrization is crucial for spiral imaging, and it seems that neither our algorithm nor reparametrization technique provides a universal answer to the issue of spiral sampling in MRI.

### 4.5.3 TSP sampling

In the same spirit of Fig. 4.3, we performed numerical experiments using a TSP trajectory (Chauffert et al., 2014a; Chauffert et al., 2013c). To perform a comparison at constant traversal time, we draw two sets of 4,500 and 45,000 "cities" in order to design a short and a long trajectory (Fig. 4.7 top row-right). The short curve is traversed with optimal reparameterization in a given time $T_{\texttt{Rep}} = 160$ ms (Fig. 4.7 middle row-right). The longer curve is parameterized at constant speed such that $T = T_{\texttt{Rep}}$, that corresponds to 25 % of the maximal speed $\gamma G_{\max}$. Then, this parameterization is projected onto $\boldsymbol{\mathcal{S}}$ (Fig. 4.7 last row).
We notice that for a fixed time, the curve obtained with our algorithm provides a larger $k$-space coverage compared to optimal reparameterization. The main reason is that TSP trajectories embody singular points that require the gradients to be set to zero for each of them. Therefore, a sampling trajectory with singular points is time consuming. The main advantage of our algorithm is that the trajectory can be smoothed around these points, which saves a lot of acquisition time. In terms of image quality, the main consequence is that our projection algorithm outperforms the reparameterization approach by 3.2 dB.

This example demonstrates that existing methods do not permit to implement TSP-based sequences in many MRI modalities (e.g., short $TE$ for a small number of "cities"), since the time to collect data can be larger than any realistic repetition time (here, the traversal time of the longer trajectory based on optimal reparameterization would require 1.1 s). In contrast, our method enables traversal of such curves in a reasonable time which can be tuned according to the image weighting ($T_1$, $T_2$ or proton density).

TABLE 4.1: Comparison between traversal time and reconstruction SNR for optimal reparameterization and projection

| | | EPI | Spiral | TSP-based |
|---|---|---|---|---|
| Resolution | | 128 | 512 | 512 |
| Optimal reparam. | $T_{\texttt{Rep}}$ (ms) | 89.6 | 42.4 | 180 |
| | SNR (dB) | 20.1 | 7.9 | 11.3 |
| Projection | $T$ (ms) | 68.9 | 200 | 180 |
| | SNR (dB) | 21.4 | 12.7 | 14.5 |

### 4.5.4 Nonlinear image reconstruction

To demonstrate the effectiveness of the proposed approach not only for gradient waveform design but also for imaging, we performed nonlinear image reconstruction as prescribed in the CS context (Candès et al., 2006a; Lustig et al., 2007). Additionally, to fully take advantage of the projection algorithm, our reconstruction scheme was non-Cartesian. Hence, we used non-uniform Fourier transforms (Keiner et al., 2009) to compute the $k$-space values out of the grid (on locations $s(i), i = 1, \ldots, n$). For comparison purposes, we started from a high resolution phantom $u$ (see Fig. 4.4) that was used to compute the sets $\mathcal{E}(u, s_{\text{Rep}})$ and $\mathcal{E}(u, s_{\text{proj}})$. The latter are given by Eq. (4.1) where $s_{\text{Rep}}$ and $s_{\text{proj}}$ denote the optimal reparameterization and projected trajectory, respectively. Next, the images were reconstructed using non-linear $\ell_1$ penalization, i.e.:

$$u^* = \arg\min_{\tilde{u}} \| \sum_{i=1}^{n} \widehat{(u - \tilde{u})}(s(i))\|_2^2 + \lambda \|\Phi \tilde{u}\|_1 \tag{4.13}$$

where $\Phi$ is a sparsifying transform (here Daubechies wavelets), $\lambda$ is a hyper-parameter, and $s$ is either $s_{\text{rep}}$ or $s_{\text{proj}}$. The minimizer of (4.13) was computed using accelerated proximal gradient descent ((Nesterov, 1983), FISTA (Beck and Teboulle, 2009b)). The image solutions ($u^*_{\texttt{Rep}}$ and $u^*_{\texttt{proj}}$) were then compared to a low resolution version of the $N \times N$ phantom where $N$ ranged from 128 to 512 to compute SNR values in Tab. 4.1. On top of this, it is worth noting that we could still improve the SNR of reconstructed images by resorting either to more redundant decompositions such as tight frames (Florescu et al., 2014) or even by learning dictionaries over which the image can be sparsely decomposed (Huang et al., 2014). However, this aspect is beyond the scope of our current proof of concept.

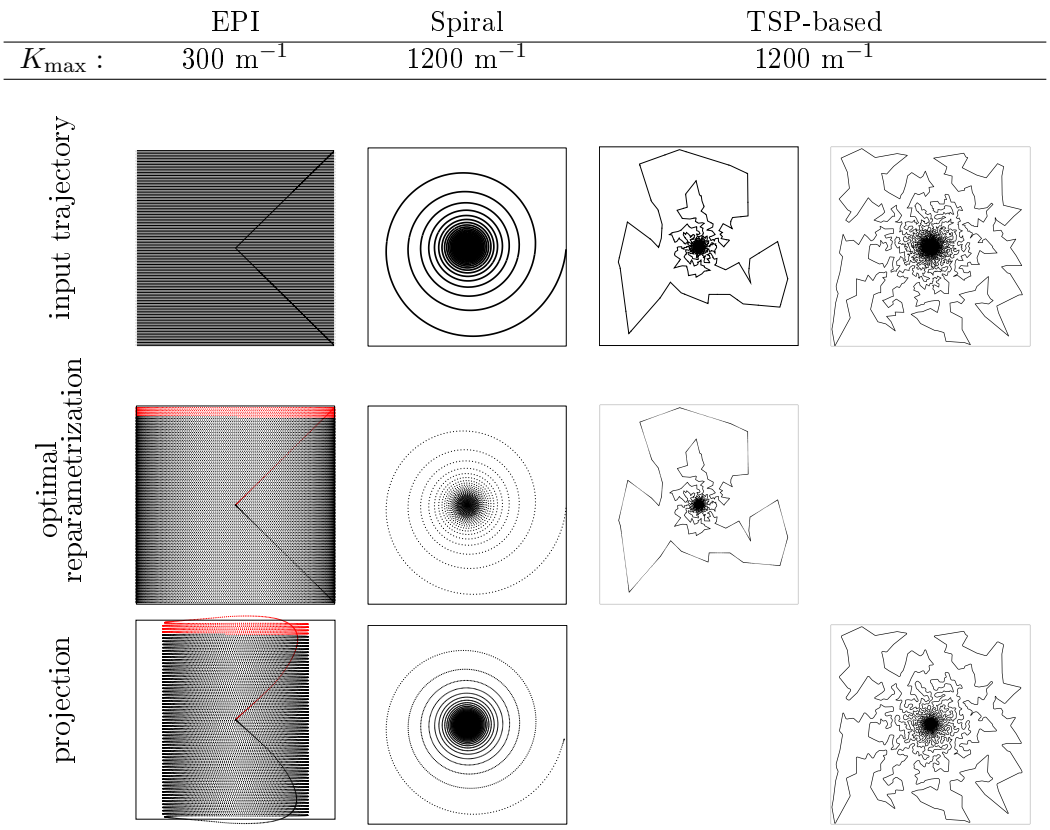


FIGURE 4.7: Representation of input trajectory, optimal reparametrization and projection for EPI, spiral and TSP-based trajectories. The frame $[-K_{\max}, K_{\max}]^2$ is depicted with various values of $K_{\max}$ that depend on the reconstruction resolution.

## 4.6 Discussion

In this paper, it has been shown that our projection algorithm has potential interests for smoothing sampling curves such as EPI or TSP-based trajectories. In this context, our algorithm delivers physically plausible trajectories while drastically reducing the traversal time and improving image quality. This is a direct consequence of its ability to project any piecewise linear initial parameterization onto admissible trajectories with different support. In applications such as functional MRI, this offers the opportunity to shorten the echo train length and then to optimally select the effective echo time so as to maximize the blood oxygenated level-dependent contrast (e.g., TE = 30 ms at 3T). Finally, our method can be used in addition with other acceleration methods such as parallel imaging (Pruessmann et al., 1999; Griswold et al., 2002) or simultaneous multi-slice imaging technique (Feinberg et al., 2010).

Beyond this context, our projection method provides a more accurate control of the sampling density as shown for variable density sampling on spirals. This has a positive

impact on image reconstruction quality at the expense of longer traversal times. Setting a fair trade-off between image quality and acquisition time is a usual concern in MRI that may depend on the application at hand (e.g., static vs dynamic imaging). Interestingly, our algorithm prescribes the acquisition time a priori what actually provides the practitioner with an effective control on such trade-off. As we illustrated on TSP sampling, this acquisition time is tightly linked to sampling accuracy with respect to the target density. Hence, our approach clearly compensates a major drawback of reparametrization methods that do not offer such control: the traversal time can be too fast hence an insufficient number of data are collected (spiral case), or too slow and not implementable (TSP-based sampling case).

Usually in MRI acquisition, a number of trajectories are interleaved to provide enough $k$-space samples. So far, we have not demonstrate the optimization of a set of interleaves except that the segmentation of the trajectory can directly enter in our global optimization problem through affine constraints if the interleaving sequence is thought of as a way of crossing the $k$-space center at evenly spaced time intervals. More generally, we can prove theoretically and practically that if the combination of two input trajectories provides a good $k$-space coverage, the combination of the two projected curves admits the same property. The theoretical argument comes from the following observation: if we consider two interleaves $c_1$ and $c_2$ and apply our projection method by searching for $s_1$ and $s_2$ from initial candidates $c_1$ and $c_2$ respectively, we actually control an upper bound of $W_2(P_{s_1}, P_{c_1}) + W_2(P_{s_2}, P_{c_2})$. Practical illustration of this property is available in our Matlab toolbox.

On the other hand, our projection method has also limitations. In particular, the projected trajectory strongly depends on the initial parameterization. As we illustrated, parameterizing a given initial curve at different speeds provides very different projected trajectories. This clearly calls for extensions that might iterate until convergence between the two key steps, namely approximating the target density and finding an admissible trajectory from this approximation (Chauffert et al., 2015a). In such generalizations, the first step can be seen as a density-consistency stage where the sampled $k$-space locations might change from one iteration to the next to fit a target density. We believe that this idea might become the most important aspect of our contribution in the future: projections are one of the most basic tools from optimization and might serve in many different contexts.

## 4.7 Conclusion

We developed an algorithm to project any parameterized curve onto the set of curves which can be implemented on actual MRI scanners. Our method is an alternative to the existing gradient waveform design based on optimal control. The major advantages are that: i) the sampling time is fixed which is crucial to adapt the proposed scheme to any MR imaging modality; ii) the sampling density is close to the target one, as required by compressed sensing theory; iii) the behavior of our algorithm is similar to the state-of-the art for smooth trajectories, but it provides shorter $k$-space coverage when the trajectory comprises numerous high curvature points, as illustrated in the TSP and EPI cases.

## Acknowledgments

This work benefited from the support of the FMJH Program Gaspard Monge in optimization and operation research, and from the support to this program from EDF. We are also grateful to our colleagues Alexandre Vignaud who carefully reread the manuscript to better target the MRI readership, and Elvis Dohmatob for his remarks.

## Appendix 1 - Density deviation, control of $W_2$-distance.

In Section 4.3.1, we aim at controlling the Wasserstein distance $W_2(P_{s^*}, \pi)$, where $\pi$ is a target fixed sampling distribution, and $P_{s^*}$ is the empirical distribution of the projected curve. We used the triangle inequlity (4.6) to bound this quantity by $W_2(P_{s^*}, P_c) + W_2(P_c, \pi)$. Here, we show that the quantity $W_2(P_c, \pi)$ can be as small as possible if $c$ is Variable Density Sampler (VDS) (Chauffert et al., 2014a). First, we define the concept of VDS, and then we provide two examples. Next, we show that if $c$ is a VDS, $W_2(P_c, \pi)$ tends to 0 as the length of $c$ tends to infinity.

### Definition of a VDS

First, we need to introduce the definition of weak convergence for measure:

**Definition 4.8.** A sequence of measures $\mu_n \in \mathcal{P}(K)$, the set of distributions defined over $K$, is said to weakly converge to $\mu$ if for any bounded continuous function $\phi$

$$\int_K \phi(x) \mathrm{d}\mu_n(x) \to \int_K \phi(x) \mathrm{d}\mu(x).$$

We use the notation $\mu_n \rightharpoonup \mu$.

According to (Chauffert et al., 2014a), a (generalized) $\pi$-VDS is a set of times $T_n$, such that $T_n \to \infty$ when $n \to \infty$, and a sequence of curves $c_{T_n} : [0, T_n] \to \mathbb{R}^d$ such that $P_{c_{T_n}} \rightharpoonup \pi$ when $n$ tends to infinity. A consequence of the definition is that the relative time spent by the curve in a part of the $k$-space is proportional to its density. Before showing that this implies that $W_2(P_{c_{T_n}}, \pi)$ tends to 0, we give two examples of VDS.

### VDS examples

We give two examples to design continuous sampling trajectories that match a given distribution. The two examples we propose provide a sequence of curves, hence a sequence of empirical measures that weakly converge to the target density.

#### Spiral sampling

The spiral-based variable density sampling is now classical in MRI (Spielman et al., 1995; Kim et al., 2003). For example, let $n \in \mathbb{R}_+$ be the number of revolutions, $r : [0,1] \mapsto \mathbb{R}^+$ a strictly increasing smooth function, and $\theta : [0,1] \to [0, 2\pi]$. Denote by $r^{-1}$ the inverse function of $r$. Define the spiral for $t \in [0, n]$ by $c_n(t) = r\left(\frac{t}{n}\right) \exp\left(i \cdot n \cdot \theta\left(\frac{t}{n}\right)\right)$ and the target distribution $\pi$ by:

$$\pi(x,y) = \begin{cases} \frac{\dot{r^{-1}}\left(\sqrt{x^2+y^2}\right)}{2\pi \int_{r(0)}^{r(1)} \dot{r^{-1}}(\rho)\rho \mathrm{d}\rho} & \text{if} \quad r(0) \leqslant \sqrt{x^2+y^2} \leqslant r(1) \\ 0 & \text{otherwise} \end{cases} \tag{4.14}$$

then $P_{c_n} \rightharpoonup \pi$ when $n$ tends to infinity.

#### Travelling Salesman-based sampling

The idea of using the shortest path amongst a set of points (the "cities") to design continuous trajectories with variable densities has been justified in (Chauffert et al., 2013a; Chauffert et al., 2014a). Let us draw $n$ $k$-space locations uniformly according to a density $q$ define over the $d$D $k$-space ($d = 2$ or 3), and join them by the shortest path (the Travelling Salesman solution). Then, denote by $c_n$ a constant-speed parameterization of this curve. Define the density:

$$\pi = \frac{q^{(d-1)/d}}{\int q^{(d-1)/d}(x)\mathrm{d}(x)}$$

Then $P_{c_n} \rightharpoonup \pi$ when the number of cities $n$ tends to infinity.

These two sampling strategies are efficient to cover the $k$-space according to target distributions, as depicted in Fig. 4.7(top row) where TSP (resp. spiral) is a VDS for distribution depicted in Fig. 4.2(a) (resp. (b)). For spiral sampling, the target distribution may be any 2D radial distribution, whereas the Travelling salesman-based sampling enable us to consider any 2D or 3D density.

### Control of $W_2$ distance

Let us now assume without loss of generality that $K = [-k_{\max}, k_{\max}]^d$.

Let us recall a central result about $W_2$ (see e.g.,(Villani, 2008)):

**Proposition 4.9.** *Let $M \subset \mathbb{R}^d$, $\mu \in \mathcal{P}(M)$ and $\mu_n$ be a sequence of $\mathcal{P}(M)$. Then, if $M$ is compact*

$$\mu_n \rightharpoonup \mu \Leftrightarrow W_2(\mu_n, \mu) \to 0$$

An immediate consequence of this proposition and of the compactness of $K$ is the following proposition:

**Proposition 4.10.** *Let $(c_{T_n})_{n\geqslant 1}$ be a $\pi$-VDS, and $\varepsilon > 0$. Then, there exists $n \geqslant 1$ such that $c_{T_n} : [0, T_n] \to K$ fulfills:*

$$W_2(P_{c_{T_n}}, \pi) \leqslant \varepsilon.$$

To sum up, Proposition 4.10 ensures that we can find an input curve which empirical distribution is as close to the target distribution $\pi$ as we want.

## Appendix 2 - Proof of Proposition 4.4

**Definition 4.11** (indicator function)**.** Let $\boldsymbol{B} \subseteq \mathbb{R}^n$. The indicator of $\boldsymbol{B}$ is denoted $\imath_{\boldsymbol{B}}$ and defined by:

$$\imath_{\boldsymbol{B}}(x) = \begin{cases} 0 & \text{if } \boldsymbol{x} \in \boldsymbol{B} \\ +\infty & \text{otherwise} \end{cases}$$

Let us now recall a classical result of convex optimization (Hiriart-Urruty and Lemaréchal, 1996, P. 195):

**Proposition 4.12.** *Let* $\boldsymbol{B}_\alpha = \{\boldsymbol{x} \in \mathbb{R}^n, \|\boldsymbol{x}\| \leqslant \alpha\}$. *Then the following identity holds:*

$$\imath_{\boldsymbol{B}_\alpha}(\boldsymbol{x}) = \sup_{\boldsymbol{y}\in\mathbb{R}^n} \langle \boldsymbol{x}, \boldsymbol{y}\rangle - \alpha\|\boldsymbol{y}\|_*.$$

Now, we can prove Proposition 4.4.

$$\begin{aligned}
&\min_{\boldsymbol{s}\in\boldsymbol{\mathcal{S}}\cap\boldsymbol{\mathcal{A}}} \frac{1}{2}\|\boldsymbol{s}-\boldsymbol{c}\|_2^2 \\
&= \min_{\boldsymbol{s}\in\boldsymbol{\mathcal{A}}} \frac{1}{2}\|\boldsymbol{s}-\boldsymbol{c}\|_2^2 + \imath_{\boldsymbol{B}_\alpha}(\dot{\mathbf{M}}\boldsymbol{s}) + \imath_{\boldsymbol{B}_\beta}(\ddot{\mathbf{M}}\boldsymbol{s}) \\
&= \min_{\boldsymbol{s}\in\boldsymbol{\mathcal{A}}} \frac{1}{2}\|\boldsymbol{s}-\boldsymbol{c}\|_2^2 + \sup_{\boldsymbol{q}_1,\boldsymbol{q}_2\in\mathbb{R}^{n\cdot d}} \langle \dot{\mathbf{M}}\boldsymbol{s}, \boldsymbol{q}_1\rangle - \alpha\|\boldsymbol{q}_1\|_* \\
&\qquad\qquad + \langle \ddot{\mathbf{M}}\boldsymbol{s}, \boldsymbol{q}_2\rangle - \beta\|\boldsymbol{q}_2\|_\star \\
&= \sup_{\boldsymbol{q}_1,\boldsymbol{q}_2\in\mathbb{R}^{n\cdot d}} \min_{\boldsymbol{s}\in\boldsymbol{\mathcal{A}}} \frac{1}{2}\|\boldsymbol{s}-\boldsymbol{c}\|_2^2 + \langle \boldsymbol{s}, \dot{\mathbf{M}}^*\boldsymbol{q}_1\rangle + \langle \boldsymbol{s}, \ddot{\mathbf{M}}^*\boldsymbol{q}_2\rangle \\
&\qquad\qquad - \alpha\|\boldsymbol{q}_1\|_* - \beta\|\boldsymbol{q}_2\|_*
\end{aligned}$$

The relationship between the primal and dual solutions reads $\boldsymbol{s}^* = \arg\min_{\boldsymbol{s}\in\boldsymbol{\mathcal{A}}} \frac{1}{2}\|\boldsymbol{s}-\boldsymbol{c}\|_2^2 + \langle \boldsymbol{s}, \dot{\mathbf{M}}^*\boldsymbol{q}_1^*\rangle + \langle \boldsymbol{s}, \ddot{\mathbf{M}}^*\boldsymbol{q}_2^*\rangle$. The sup and the min can be interverted at the third line, due to standard theorems in convex analysis (see e.g. (Rockafellar, 1997, Theorem 31.3)).

## Appendix 3 - Proof of Propositions 4.5

To show Proposition 4.5, first remark that

$$\begin{aligned}
&\arg\min_{\boldsymbol{s}\in\boldsymbol{\mathcal{A}}} \langle \dot{\mathbf{M}}\boldsymbol{s}, \boldsymbol{q}_1\rangle + \langle \ddot{\mathbf{M}}\boldsymbol{s}, \boldsymbol{q}_2\rangle + \frac{1}{2}\|\boldsymbol{s}-\boldsymbol{c}\|_2^2 \\
&= \arg\min_{\boldsymbol{s}\in\boldsymbol{\mathcal{A}}} \frac{1}{2}\|\boldsymbol{s} - (\boldsymbol{c} - \dot{\mathbf{M}}\boldsymbol{q}_1 - \ddot{\mathbf{M}}^*\boldsymbol{q}_2)\|_2^2.
\end{aligned}$$

Therefore, $\boldsymbol{s}^*(\boldsymbol{q}_1, \boldsymbol{q}_2)$ is the orthogonal projection of $\boldsymbol{z} = \boldsymbol{c} - \dot{\mathbf{M}}\boldsymbol{q}_1 - \ddot{\mathbf{M}}^*\boldsymbol{q}_2$ onto $\boldsymbol{\mathcal{A}}$. Since $\boldsymbol{\mathcal{A}}$ is not empty, $\boldsymbol{A}\boldsymbol{A}^+\boldsymbol{v} = \boldsymbol{v}$, and the set $\boldsymbol{\mathcal{A}} = \{\boldsymbol{s} \in \mathbb{R}^{n\cdot d}, \boldsymbol{A}\boldsymbol{s} = \boldsymbol{v}\}$ can be decomposed as

$$\boldsymbol{\mathcal{A}} = \boldsymbol{A}^+\boldsymbol{v} + \ker(\boldsymbol{A}).$$

The vector $\boldsymbol{z} - \boldsymbol{s}^*(\boldsymbol{q}_1, \boldsymbol{q}_2)$ is orthogonal to $\boldsymbol{\mathcal{A}}$, it therefore belongs to $\ker(\boldsymbol{A})^\perp = \mathrm{im}(\boldsymbol{A}^*)$. Hence $\boldsymbol{s}^*(\boldsymbol{q}_1, \boldsymbol{q}_2) = \boldsymbol{z} + \boldsymbol{A}^*\boldsymbol{\lambda}$ for some $\boldsymbol{\lambda}$ such that:

$$\boldsymbol{A}(\boldsymbol{z} + \boldsymbol{A}^*\boldsymbol{\lambda}) = \boldsymbol{v}.$$

This leads to $\boldsymbol{\lambda} = (\boldsymbol{A}\boldsymbol{A}^*)^{-1}(\boldsymbol{v} - \boldsymbol{A}\boldsymbol{z})$. We finally get

$$\boldsymbol{s}^*(\boldsymbol{q}_1, \boldsymbol{q}_2) = \boldsymbol{z} + \boldsymbol{A}^*(\boldsymbol{A}\boldsymbol{A}^*)^{-1}(\boldsymbol{v} - \boldsymbol{A}\boldsymbol{z}),$$

ending the proof.

## Appendix 4 - Proof of theorem 4.7.

Let us first recall that the relative interior of a convex set $C$ $\mathrm{ri}(C)$ is the interior of $C$ relative to the affine hull of $C$ (Hiriart-Urruty and Lemaréchal, 1996). The analysis proposed to prove Theorem 4.7 closely follows ideas proposed in (Weiss et al., 2009; Boyer et al., 2014; Beck and Teboulle, 2009b; Beck and Teboulle, 2014). We will need two results. The first one is a duality result from (Boyer et al., 2014).

**Proposition 4.13.** *Let $f : \mathbb{R}^m \to \mathbb{R} \cup \{\infty\}$ and $g : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ denote two closed convex functions, and $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ denote a matrix. Assume that $g$ is $\sigma$-strongly convex (Hiriart-Urruty and Lemaréchal, 1996) and that $\boldsymbol{A}\mathrm{ri}(\mathrm{dom}(f)) \cap \mathrm{ri}(\mathrm{dom}(g)) \neq \emptyset$.*

*Let $p(\boldsymbol{x}) = f(\boldsymbol{A}\boldsymbol{x}) + g(\boldsymbol{x})$ and $d(\boldsymbol{y}) = -g^*(\boldsymbol{A}^*\boldsymbol{y}) - f^*(\boldsymbol{y})$. Let $\boldsymbol{x}^*$ be the unique minimizer of $p$ and $\boldsymbol{y}^*$ be any minimizer of $d$.*

*Then $g^*$ is differentiable with $\frac{1}{\sigma}$ Lipschitz-continuous gradient. Moreover, by letting $\boldsymbol{x}(\boldsymbol{y}) = \nabla g^*(-\boldsymbol{A}^*\boldsymbol{y})$:*

$$\|\boldsymbol{x}(\boldsymbol{y}) - \boldsymbol{x}^*\|_2^2 \leq \frac{2}{\sigma}(d(\boldsymbol{y}) - d(\boldsymbol{y}^*)).$$

The second ingredient is the standard convergence rate for accelerated proximal gradient descents given in (Beck and Teboulle, 2009b, Theorem. 4.4).

**Proposition 4.14.** *Under the same assumptions as Proposition 4.13, consider Algorithm 2.*

**Algorithm 2:** Accelerated proximal gradient descent

**Input**: $\boldsymbol{q}_0 \in \mathrm{ri}(\mathrm{dom}(f^*)) \cap \boldsymbol{A}\mathrm{ri}(\mathrm{dom}(g^*))$ and $n_{it}$
**Initialize** Set $\ell = 1/L$, with $L = \frac{|||\boldsymbol{A}|||^2}{\sigma}$.
Set $\boldsymbol{y}_0 = \boldsymbol{q}_0$. **for** $k = 1 \dots n_{it}$ **do**
  $\boldsymbol{q}^{(k)} = \mathbf{prox}_{\ell f^*}(\boldsymbol{y}^{(k-1)} + \ell \boldsymbol{A}\nabla g^*(-\boldsymbol{A}^*\boldsymbol{y}^{(k-1)}))$
  $\boldsymbol{y}^{(k)} = \boldsymbol{q}^{(k)} + \frac{k-1}{k+2}(\boldsymbol{q}^{(k)} - \boldsymbol{q}^{(k-1)})$

*Then $\|\boldsymbol{y}^{(n_{it})} - \boldsymbol{y}^*\|_2^2 = \mathcal{O}\left(\frac{|||\boldsymbol{A}|||^2}{\sigma \cdot n_{it}^2}\right)$.*

To conclude, it suffices to set $g(\boldsymbol{s}) = \frac{1}{2}\|\boldsymbol{s} - \boldsymbol{c}\|_2^2$, $f(\boldsymbol{q}_1, \boldsymbol{q}_2) = \imath_{\boldsymbol{B}_\alpha}(\boldsymbol{q}_1) + \imath_{\boldsymbol{B}_\alpha}(\boldsymbol{q}_2)$ and $\boldsymbol{A} = \begin{pmatrix} \dot{\mathbf{M}} \\ \ddot{\mathbf{M}} \end{pmatrix}$. By doing so, the projection problem rewrites $\min_{\boldsymbol{s}\in\mathbb{R}^{n\dot{d}}} p(\boldsymbol{s}) = f(A\boldsymbol{s}) + g(\boldsymbol{s})$. Its dual problem (4.8) can be rewritten more compactly as $\min_{\boldsymbol{q}=(\boldsymbol{q}_1,\boldsymbol{q}_2)\in\mathbb{R}^{n\dot{d}}\times\mathbb{R}^{n\dot{d}}} d(\boldsymbol{q}) = g^*(-\boldsymbol{A}^*\boldsymbol{q}) + f^*(\boldsymbol{q})$. Note that function $g$ is 1-strongly convex. Therefore, Algorithm 2 can be used to minimize $d$, ensuring a convergence rate in $\mathcal{O}\left(\frac{L}{k^2}\right)$ on the function values $d(\boldsymbol{y}^{(k)})$, where $L = |||\boldsymbol{A}|||^2$. It then suffices to use Proposition 4.13 to obtain a convergence rate on the distance to the solution $\|\boldsymbol{s}^{(k)} - \boldsymbol{s}^*\|_2^2$. This ends the proof of Theorem 4.7.

## 4.8 Additionnal simulations on 3D angiography

This section was published in the proceedings (Chauffert et al., 2015b) and shows the behaviour of the projection algorithm on 3D TSP-based sampling trajectories and its application to angiography.

In this part, we compare the time to traverse $k$-space along different trajectories using gradients computed either by the standard optimal control approach or by our proposed projection algorithm. For comparison between sampling schemes, we work on *retrospective* CS, meaning that a full dataset has been acquired, and then a posteriori downsampling is performed. We compare the reconstruction results in terms of peak signal-to-noise ratio (PSNR) with respect to the acquisition time and to the "acceleration factor"[6] $r$.

### 4.8.1 Experimental framework

**Data acquisition.** The initial experimental setup aimed at observing blood vessels of living mice using an intraveinous injection of an iron oxide-based contrast agent (Magnetovibrio Blakemorei MV1). Because of natural elimination, it is necessary to speed up acquisition to improve contrast and make easier post-processing such as angiography. The experiments have been performed on a 17.2T preclinical scanner which physical rotation-invariant constraints are, for all $t \in [0, T]$:

$$\|g(t)\| \leqslant 1 \text{ T.m}^{-1} \qquad \text{and} \qquad \|\dot{g}(t)\| \leqslant 5.3 \text{ T.m}^{-1}.\text{ms}^{-1}.$$

A FLASH sequence (Fast Low Angle SHot) has been used to reveal the $T_2^*$ contrast induced by the injection of the contrast agent (TE/TR = 8/680 ms). The sequence was repeated 12 times to improve the signal-to-noise ratio (SNR), leading to a total acquistion

[6] $r$ quantifies the reduction of the number of measurements $m$. If the $k$-space is a grid of $N$ pixels $r := N/m$ is commonly used in CS-MRI.

time of 30 minutes to acquire the $k$-space slice by slice. The spatial resolution achieved is $90\times90\times180$ $\mu$m$^3$.

**Hypothesis.** The aim of this experiment is to prove that one can expect a large acquisition time reduction using partial $k$-space measurements. The time to traverse a sampling curve is computed satisfying the gradient constraints. To achieve a fair comparison, let us mention the additional hypothesis that our acquisitions are single-shot, meaning that the partial $k$-space is acquired after a single RF pulse. We did not take echo and repetition times into account to ensure the recovery of a $T_2^*$-weigthed image. We only compare the time to traverse a curve using the gradients with their maximal intensity. We assume that there is no error on the $k$-space sample locations. In practice we have to measure the three magnetic field gradients that are actually played out by the scanner to correct the trajectory and avoid distortions. We shall work on a discrete cartesian $k$-space, and consider that a sample is measured if the sampling trajectory crosses the corresponding cell of the $k$-space grid. Using this hypothesis, the estimated time to visit the 2D $k$-space is 110 ms.

**Strategy.** We used the TPS-based sampling method (Appendix 1) as input of our projection algorithm (see Fig. 4.8(b)), since it is a way of designing sampling trajectories that match any sampling density $\pi$. The latter is central in CS-MRI since it impacts the number of required measurements (Adcock et al., 2013; Krahmer and Ward, 2014; Chauffert et al., 2014a). To compare our projection method to existing reparameterization, the proposed sampling strategy is:
**(i)** Sample deterministically the $k$-space center as adviced in (Adcock et al., 2013; Chauffert et al., 2013b; Chauffert et al., 2014a), using an EPI sequence (see Fig. 4.8(a)). The scanning time can be estimated to 12 ms in 2D using optimal control.
**(ii)** Select a density $\pi$ proportional to $1/|k|^2$ as mentioned in (Krahmer and Ward, 2014; Chauffert et al., 2014a). Draw independently points according to $\pi^{\frac{d-1}{d}}$ and join them by the shortest path to form a $\pi$-VDS (Appendix 1).
**(iii)** Parameterize the TSP path at *constant speed* and project this parameterization onto the set of gradient constraints, or **(iii bis)** Parameterize the TSP path using optimal control (the exact solution can be computed explicitely).
**(iv)** Form the sampling curve, define a set $\Omega$ of the selected samples, mask the $k$-space with $\Omega$, and reconstruct an image using $\ell_1$ minimization of the constrained problem. Let $\mathbf{F}^*$ denote the $d$-dimensional discrete Fourier transform and $\mathbf{F}_\Omega^*$ the matrix composed of the lines corresponding to $\Omega$. Denote also by $\mathbf{\Phi}$ an inverse $d$-dimensional wavelet transform (here a Symmlet transform). Then the reconstructed image is the solution of the

problem:

$$x^* = \underset{y=\mathbf{F}_\Omega^* x}{\text{Argmin}} \left\| \mathbf{\Phi}^{-1} x \right\|_1 \tag{4.15}$$

An approximation of $x^*$ is computed using Douglas-Rachford algorithm (Combettes and Pesquet, 2011). Solving the penalized form associated with (4.15) might be addressed by competing algorithms (ADMM, 3MG); see (Florescu et al., 2014) for a recent comparison. The reconstruction results could be improved by resorting to non-Cartesian reconstruction (Keiner et al., 2009), which would avoid the approximation related to the projection onto the $k$-space grid.

### 4.8.2 Results

#### 4.8.2.1 2D reconstructions

In this experiment, we considered a 2D $k$-space ($d = 2$) corresponding to an axial slice. We considered five sampling strategies, depicted in Fig. 4.8(first row): a classical EPI coverage used as reference (a); a TSP-based sampling trajectory parameterized using optimal control (b); two projected TSP-based trajectories, one with the same number of samples collected as in (b) ($r = 11.2$) (c) and the other with the same scanning time as in (b) (62 ms) (d); a variable density spiral trajectory for comparison purpose in terms of time and sampling ratio (e).

As expected, the reconstruction results shown in Fig. 4.8(g,h) are really close, since the number of collected samples is the same, and the sampling densities are similar. However, in this comparison the gain in traversal time is significant (one half). In contrast, the longer and smoothed TSP depicted in Fig. 4.8(d) allows us to improve image reconstruction (1 dB gain) as illustrated by Fig. 4.8(i) while keeping the same acquisition time as in Fig. 4.8(b). For comparison purposes, we implemented spiral acquisition which consists of replacing steps (ii)-(iii) in the above mentioned sampling strategy by a spiral with density proportional to $1/|k|^2$, projected onto the set of constraints. This strategy doubles the acquisition time (118 ms compared to 62 ms) whereas the acceleration factor was larger ($r = 7.5$ vs. $r = 6.6$). In this experimental context (regridding and variable density spiral), the spiral is not appealing compared to EPI acquisition, since it is time consuming and degrades the image quality.

In each of these reconstructions, the major vessels can be recovered, although the smallest ones can only be seen for $r < 8$. Finally, the best compromise between acquisition time and reconstruction quality is achieved using the specific combination of TSP-based sampling and our projection algorithm onto the set of constraints shown in Fig. 4.8(d).

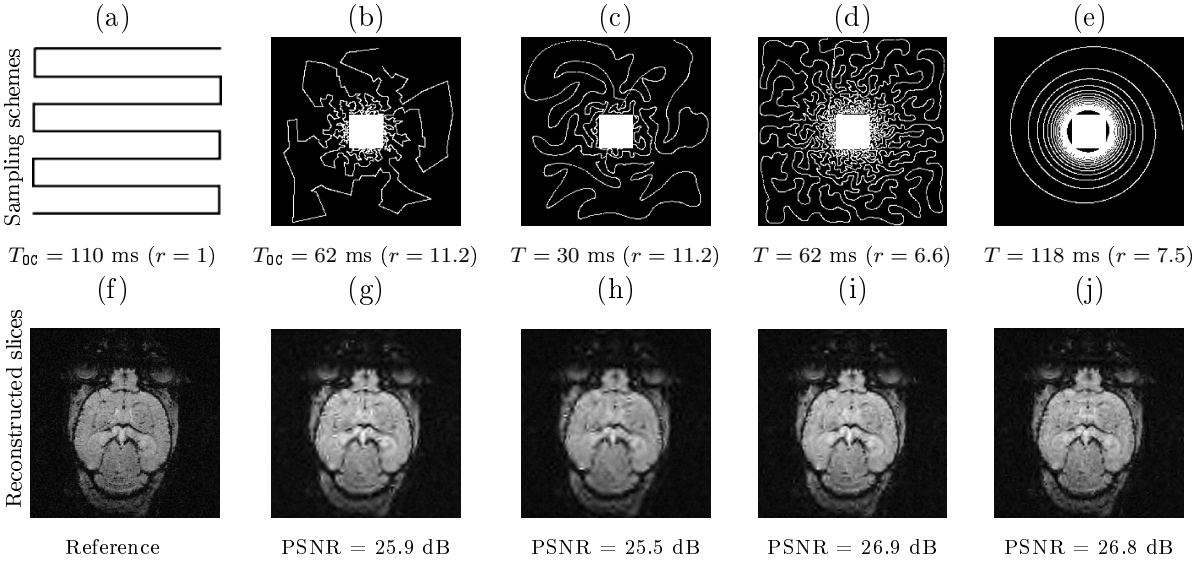


FIGURE 4.8: Full $k$-space acquisition with an EPI sequence (a) and corresponding reference image (f). Comparison between an exact parameterization of the TSP trajectory (b) and projection from TSP trajectory onto the set of constraints (c),(d). In experiments (b,c), the number of measured locations is fixed to 9% ($r = 11.2$), whereas in (b,d), the time to traverse the curve is fixed to 62 ms. (e): Spiral trajectory with full acquisition of the $k$-space center. (g-j): Reconstructed images corresponding to sampling strategies (b-e) by solving Eq. (4.15).

#### 4.8.2.2 3D angiography

Using the same method as in 2D, namely TSP-sampling and projection onto the set of constraints, we reconstructed volumes from 3D $k$-space. In order to estimate the quality of the reconstructions, we compared the angiograms computed from the 3D images using Frangi filtering (Frangi et al., 1998). The results are shown in Fig. 4.9 for acceleration factors $r = 7.3$ (Fig. 4.9(b,e)) and $r = 17.4$ (Fig. 4.9(c,f)) and compared to the angiogram computed from the whole data.

Using the strategy described in Part 4.8.1 the time to traverse $k$-space would be 3.53 s (full acquisition), 3.15 s ($r = 7.2$) and 0.88 s ($r = 17$). The main drawback of TSP-based sampling schemes is that the time reduction is not directly proportional to $r$, in contrast to classical 2D downsampling and reading out along the third dimension. Nevertheless, if the number of measurements is fixed, the TSP-based approach leads to more accurate reconstruction results since the sampling scheme may fit any density (see Appendix 1). Angiograms shown in Fig. 4.9 illustrate that one can reduce the travel time in the $k$-space and still observe accurate microvascular structure. If $r = 7.3$, time reduction is minor (about 10% less), but the computed angiogram is almost the same as the one obtained with a complete $k$-space. It is interesting to notice that with a higher acceleration factor ($r = 17.4$), the acquisition time is reduced by 75%, but the computed angiogram remains of good quality. The angiogram appears a bit noisier, especially

in the pre-injection setting (Fig. 4.9(c)), but the post-injection image allows recovering Willis polygon and most of the major vessels of the mouse brain (Fig. 4.9(f)).

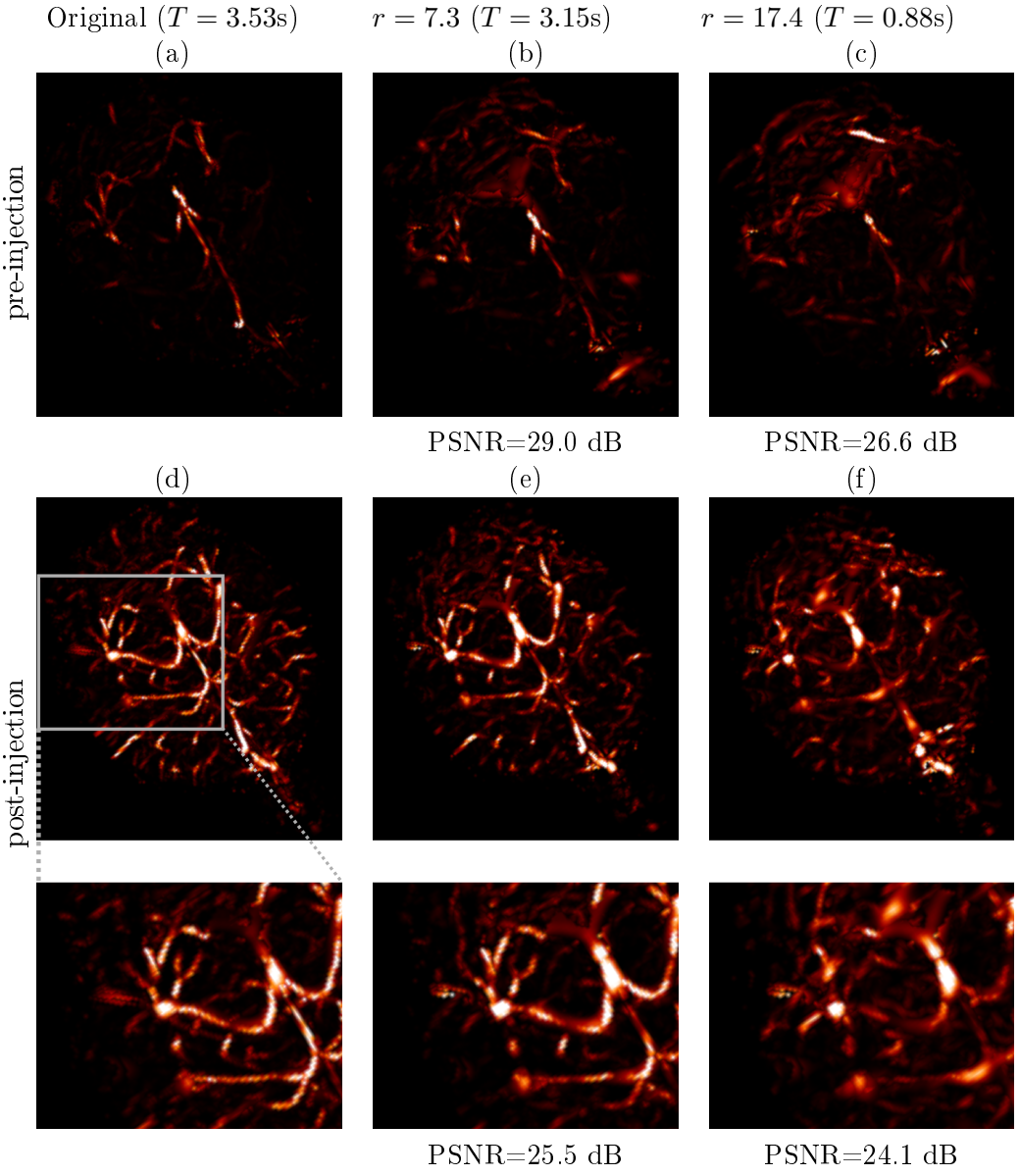


FIGURE 4.9: Angiograms computed from full $k$-space pre-(a) and post-(d) injection data. Angiograms computed from pre-(resp., post-) injection data for decimated $k$-space with $r=7.3$ (b) and $r=17.4$ (c) (resp., (e) and (f)).

# Chapter 5

# A projection method on measures sets

This chapter is based on (Chauffert et al., 2015a).

We consider the problem of projecting a probability measure $\pi$ on a set $\mathcal{M}_N$ of Radon measures. The projection is defined as a solution of the following variational problem:

$$\inf_{\mu \in \mathcal{M}_N} \|h \star (\mu - \pi)\|_2^2,$$

where $h \in L^2(\Omega)$ is a kernel, $\Omega \subset \mathbb{R}^d$ and $\star$ denotes the convolution operator. To motivate and illustrate our study, we show that this problem arises naturally in various practical image rendering problems such as stippling (representing an image with $N$ dots) or continuous line drawing (representing an image with a continuous line). We provide a necessary and sufficient condition on the sequence $(\mathcal{M}_N)_{N \in \mathbb{N}}$ that ensures weak convergence of the projections $(\mu_N^*)_{N \in \mathbb{N}}$ to $\pi$. We then provide a numerical algorithm to solve a discretized version of the problem and show several illustrations related to computer-assisted synthesis of artistic paintings/drawings.

## 5.1 Introduction

Digital Halftoning consists of representing a grayscale image with only black and white tones (Ulichney, 1987). For example, a grayscale image can be approximated by a variable distribution of black dots with over a white background. This technique, called stippling, is the cornerstone of most printing digital inkjet devices. A stippling result is displayed in Figure 5.1b. The lion in Figure 5.1a can be recognized from the dotted image shown in Figure 5.1b. This is somehow surprising since the differences between the pixel values

of the two images are far from fzero. One way to explain this phenomenon is to invoke the multiresolution feature of the human visual system (Daugman, 1980; Pappas and Neuhoff, 1999). Figures 5.1c and 5.1d are blurred versions of Figures 5.1a and 5.1b respectively. These blurred images correspond to low-pass versions of the original ones and are nearly impossible to distinguish.

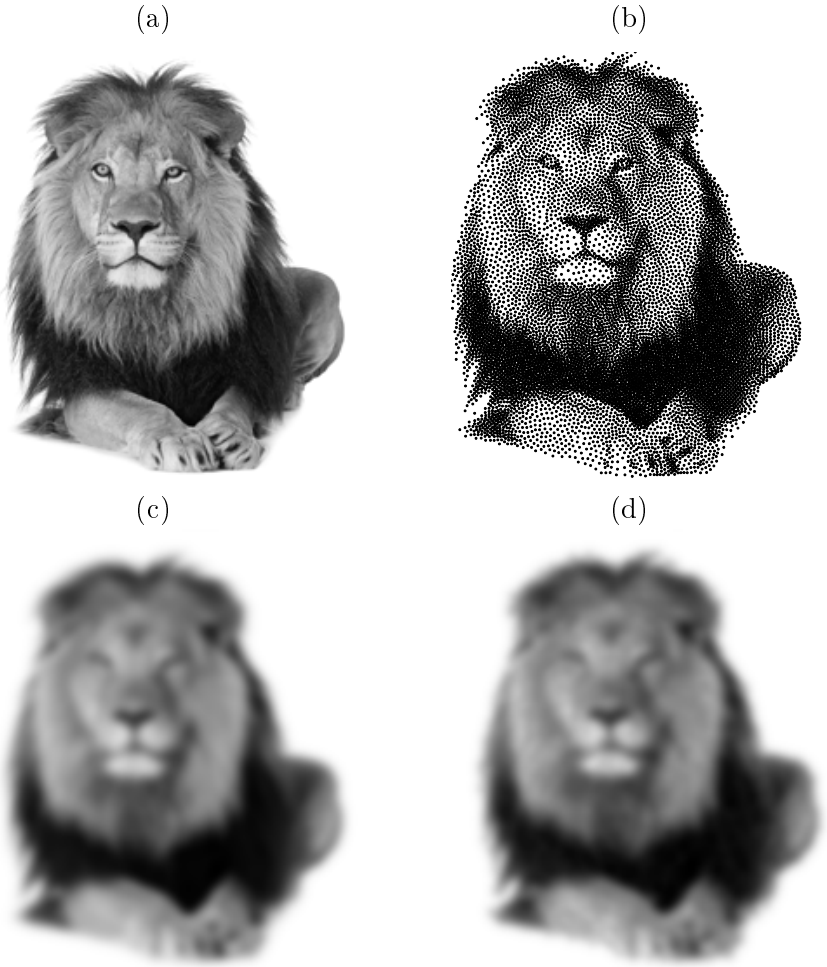


FIGURE 5.1: Explanation of the stippling phenomenon. Images (a) and (b) are similar while the norm of their difference is large. Figures (c) and (d) are obtained by convolving (a) and (b) with a Gaussian of variance equal to 3 pixels. After convolution, the images cannot be distinguished.

Assuming that the dots correspond to Dirac masses, this experiment suggests placing the dots at locations $p_1, \ldots, p_N$ corresponding to the minimizer of the following variational problem:

$$\min_{(p_1,\ldots,p_N)\in\Omega^N} \left\| h \star \left( \pi - \frac{1}{N} \sum_{i=1}^{N} \delta_{p_i} \right) \right\|_2^2 \tag{5.1}$$

where $\Omega \subset \mathbb{R}^2$ denotes the image domain, $\delta_{p_i}$ denotes the Dirac measure at point $p_i \in \mathbb{R}^2$, $\pi$ denotes the target probability measure (the lion) and $h$ is a convolution kernel that

should depend on the point spread function of the human visual system. By letting

$$\mathcal{M}(\Omega^N) = \left\{ \mu = \frac{1}{N} \sum_{i=1}^{N} \delta_{p_i},\ (p_i)_{1 \leq i \leq N} \in \Omega^N \right\} \tag{5.2}$$

denote the set of $N$-point measures, problem (5.1) rereads as a projection problem:

$$\min_{\mu \in \mathcal{M}(\Omega^N)} \|h \star (\pi - \mu)\|_2^2 . \tag{5.3}$$

This variational problem is a prototypical example that motivates our study. As explained later, it is intimately related to recent works on image halftoning by means of attraction-repulsion potentials proposed in (Schmaltz et al., 2010; Teuber et al., 2011; Gwosdek et al., 2014). In references (Gräf et al., 2012; Fornasier et al., 2013; Fornasier and Hütter, 2013) this principle is shown to have far reaching applications ranging from numerical integration, quantum physics, economics (optimal location of service centers) or biology (optimal population distributions).

In this paper, we extend this variational problem by replacing $\mathcal{M}(\Omega^N)$ with an arbitrary set of measures denoted $\mathcal{M}_N$. In other words, we want to approximate a given measure $\pi$ by another measure in the set $\mathcal{M}_N$. We develop an algorithm to perform this projection in a general setting.

To motivate this extension, we consider a practical problem: how to perform continuous line drawing with a computer? Continuous line drawing is a starting course in all art cursus. It consists of drawing a picture without ever lifting the paintbrush from the page. Figure 5.2 shows two drawings obtained with this technique. Apart from teaching, it is used in marketing, quilting designs, steel wire sculptures, connect the dot puzzles,... A few algorithms were already proposed in (Li and Mould, 2014; Xu and Kaplan, 2007; Kaplan et al., 2005; Bosch and Herman, 2004; Wong and Takahashi, 2011). We propose an original solution which consists of setting $\mathcal{M}_N$ as a space of pushforward measures associated with sets of parameterized curves.

Apart from the two rendering applications discussed in this paper, this paper has potential for diverse applications in fields such as imaging, finance, biology,...

The remaining of this paper is structured as follows. We first describe the notation and some preliminary remarks in Section 5.2. We propose a mathematical analysis of the problem for generic sequences of measures spaces $(\mathcal{M}_N)_{N \in \mathbb{N}}$ in Section 5.3. In particular, we give conditions on $h$ ensuring that the mapping $\mu \mapsto \|h \star \mu\|_2$ defines a norm on the space of signed measures and provide necessary and sufficient conditions on the sequence $(\mathcal{M}_N)_{N \in \mathbb{N}}$ ensuring consistency of the projection problem. We propose a generic numerical algorithm in Section 5.4 and derive some of its theoretical guarantees. In

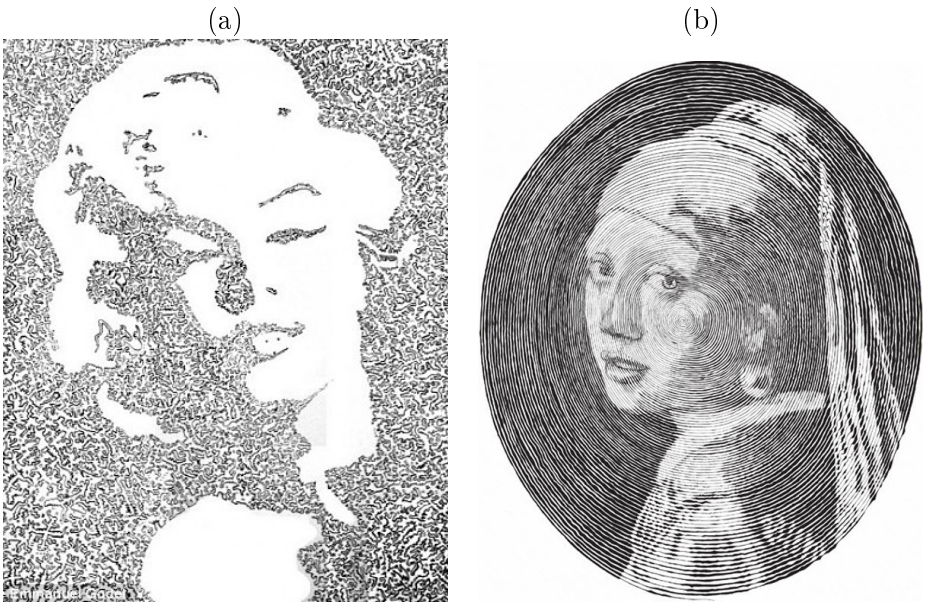


FIGURE 5.2: Two examples of continuous line drawing. (a) A sketch of Marylin Monroe by Pierre Emmanuel Godet http://pagazine.com/ using a continuous line. A close inspection reveals that the line represents objects and characters. (b) Meisje met de Parel, Vermeer 1665, represented using a spiral with variable width. Realized by Chan Hwee Chong http://www.behance.net/Hweechong.

Section 5.5, we study the particular problem of continuous line drawing from a mathematical perspective. Finally, we present some results in image rendering problems in Section 5.6.

## 5.2 Notation and preliminaries

In this paper, we work on the measurable space $(\Omega, \Sigma)$, where $\Omega = \mathbb{T}^d$ denotes the torus $\mathbb{T}^d = \mathbb{R}^d/\mathbb{Z}^d$. An extension to other spaces such as $\mathbb{R}^d$ or $[0,1]^d$ is feasible but requires slight adaptations. Since drawing on a donut is impractical, we will set $\Omega = [0,1]^d$ in the numerical experiments.

The space of continuous functions on $\Omega$ is denoted $\mathcal{C}(\Omega)$. The Sobolev space $(W^{m,p}([0,T]))^d$, where $m \in \mathbb{N}$, is the Banach space of $d$ dimensional curves in $\Omega$ with derivatives up to the $m$-th order in $L^p([0,T])$. Let $\mathcal{M}_\Delta$ denote the space of probability measures on $\Omega$, i.e. the space of nonnegative Radon measures $p$ on $\Omega$ such that $p(\Omega) = 1$. Throughout the paper $\pi \in \mathcal{M}_\Delta$ will denote a *target measure*. Let $\mathcal{M}$ denote the space of signed measures on $\Omega$ with bounded total variation, that is $\mu = \mu_+ - \mu_-$ where $\mu_+$ and $\mu_-$ are two finite nonnegative Radon measures and $\|\mu\|_{TV} = \mu_+(\Omega) + \mu_-(\Omega) < \infty$.

Let $h : \Omega \to \mathbb{R}$ denote a continuous function. Let $\mu \in \mathcal{M}$ denote an arbitrary finite signed measure. The convolution product between $h$ and $\mu$ is defined for all $x \in \Omega$ by:

$$\begin{aligned} \mu \star h(x) &\coloneqq \int_{\Omega} h(x-y) d\mu(y) \\ &= \mu(h(x-\cdot)) \end{aligned} \tag{5.4}$$

In the Fourier space, the convolution (5.4) translates to, for all $\xi \in \mathbb{Z}^d$ (see e.g., (Katznelson, 1968)):

$$\widehat{\mu \star h}(\xi) = \hat{\mu}(\xi)\hat{h}(\xi),$$

where $\hat{\mu}$ is the Fourier-Stieltjes series of $\mu$. The Fourier-Stieltjes series coefficients are defined for all $\xi \in \mathbb{Z}^d$ by:

$$\hat{\mu}(\xi) \coloneqq \int_{\Omega} e^{-2i\pi\langle \xi, x\rangle}\, d\mu(x).$$

We recall the Parseval formula:

$$\int_{\Omega} |h(x)|^2\, dx = \sum_{\xi \in \mathbb{Z}^d} \left|\hat{h}(\xi)\right|^2.$$

Let $J : \mathbb{R}^n \to \mathbb{R}$ denote a function and $\partial J$ its limiting-subdifferential (or simply subdifferential) (Mordukhovich, 2006; Attouch et al., 2013). Let $C \subseteq \mathbb{R}^n$ denote a closed subset. The indicator function of $C$ is denoted $i_C$ and defined by

$$i_C(x) = \begin{cases} 0 & \text{if } x \in C, \\ +\infty & \text{otherwise.} \end{cases}$$

The set of projections of a point $x_0 \in \mathbb{R}^n$ on $C$ is denoted $P_C(x_0)$ and defined by

$$P_C(x_0) = \operatorname*{Arg\,min}_{x \in C} \|x - x_0\|_2^2.$$

The notation Arg min stands for the whole set of minimizers while arg min denotes one of the minimizers. Note that $P_C$ is generally a point-to-set mapping except if $C$ is convex closed, since the projection on a closed convex set is unique. The normal cone at $x \in \mathbb{R}^n$ is denoted $N_C(x)$. It is defined as the limiting-subdifferential of $i_C$ at $x$. A critical point of the function $J + i_C$ is a point $x^*$ that satisfies $0 \in \partial J(x^*) + N_C(x^*)$. This condition is necessary (but not sufficient) for $x^*$ to be a local minimizer of $J + i_C$.

## 5.3 Mathematical analysis

Let

$$\mathcal{N}_h(\mu) \coloneqq \|h \star \mu\|_2. \tag{5.5}$$

In this section, we study some basic properties of the following projection problem:

$$\min_{\mu \in \mathcal{M}_N} \mathcal{N}_h(\pi - \mu), \tag{5.6}$$

where $(\mathcal{M}_N)_{N \in \mathbb{N}}$ denotes an arbitrary sequence of measures sets in $\mathcal{M}_\Delta$.

### 5.3.1 Norm properties

We first study the properties of $\mathcal{N}_h$ on the space $\mathcal{M}$ of signed measures with bounded total variation. The following proposition shows that it is well defined provided that $h \in \mathcal{C}(\Omega)$.

**Proposition 5.1.** *Let $h \in \mathcal{C}(\Omega)$ and $\mu \in \mathcal{M}$. Then $h \star \mu \in L^2(\Omega)$.*

*Proof.* It suffices to remark that $\forall x \in \Omega$, $|h \star \mu(x)| \leq \|\mu\|_{TV}\|h\|_\infty < +\infty$. Therefore, $h \star \mu \in L^\infty(\Omega)$. Since $\Omega$ is bounded, $h \in L^\infty(\Omega)$ implies that $h \in L^2(\Omega)$. □

*Remark* 5.2. In fact, the result holds true for weaker hypotheses on $h$. If $h \in \mathcal{L}^\infty(\Omega)$, the set of bounded Borel measurable functions, $h \star \mu \in L^2(\Omega)$ since

$$\forall x \in \Omega,\ |h \star \mu(x)| \leq \|\mu\|_{TV} \left( \sup_{x \in \Omega} |h(x)| \right) < +\infty.$$

Note that the $L^\infty$-norm is defined with an ess sup while we used a sup in the above expression. We stick to $h \in \mathcal{C}(\Omega)$ since this hypothesis is more usual when working with Radon measures.

The following proposition gives a necessary and sufficient condition on $h$ ensuring that $\mathcal{N}_h$ defines a norm on $\mathcal{M}$.

**Proposition 5.3.** *Let $h \in \mathcal{C}(\Omega)$. The mapping $\mathcal{N}_h$ defines a norm on $\mathcal{M}$ if and only if all Fourier series coefficients $\hat{h}(\xi)$ are nonzero.*

*Proof.* Let us assume that $\hat{h}(\xi) \neq 0$, $\forall \xi \in \mathbb{Z}^d$. The triangle inequality and absolute homogeneity hold trivially. Let us show that $\mu \neq 0 \Rightarrow \mathcal{N}_h(\mu) \neq 0$. The Fourier series of a nonzero signed measure $\mu$ is nonzero, so that there is $\xi \in \mathbb{Z}^d$ such that $\hat{\mu}(\xi) \neq 0$. According to our hypothesis $\hat{h}(\xi) \neq 0$, hence $\widehat{\mu \star h}(\xi) \neq 0$ and $\mathcal{N}_h(\mu) \neq 0$.

On the contrary, if there exists $\xi_0 \in \mathbb{Z}^d$ such that $\hat{h}(\xi_0) = 0$. The non-zero measure defined through its Fourier series by

$$\hat{\mu}(\xi) = \begin{cases} 1 & \text{if } \xi = \xi_0 \\ 0 & \text{otherwise} \end{cases}$$

satisfies $\mathcal{N}_h(\mu) = 0$ and belongs to $\mathcal{M}$. □

From now on, owing to Proposition 5.3, we will systematically assume - sometimes without mentioning - that $h \in \mathcal{C}(\Omega)$ and that $\hat{h}(\xi) \neq 0$, $\forall \xi \in \mathbb{Z}^d$. Finally, we show that $\mathcal{N}_h$ induces the weak topology on $\mathcal{M}$. Let us first recall the definition of weak convergence.

**Definition 5.4.** A sequence of measures $(\mu_N)_{N\in\mathbb{N}}$ is said to weakly converge to $\mu \in \mathcal{M}$, if

$$\lim_{N\to\infty} \int_\Omega f(x) d\mu_N(x) = \int_\Omega f(x) d\mu(x)$$

for all continuous functions $f : \Omega \to \mathbb{R}$. The shorthand notation for weak convergence is

$$\mu_N \underset{N\to\infty}{\rightharpoonup} \mu.$$

**Proposition 5.5.** *Assume that $h \in \mathcal{C}(\Omega)$ and that $\hat{h}(\xi) \neq 0$, $\forall \xi \in \mathbb{Z}^d$. Then for all sequences $(\mu_N)_{N\in\mathbb{N}}$ in $\mathcal{M}$ satisfying $\|\mu_N\|_{TV} \leq M < +\infty$, $\forall N \in \mathbb{N}$,*

$$\lim_{N\to\infty} \mathcal{N}_h(\mu_N) = 0 \quad \Leftrightarrow \quad \mu_N \underset{N\to\infty}{\rightharpoonup} 0.$$

*Proof.* Let $(\mu_N)_{N\in\mathbb{N}}$ be a sequence of signed measures in $\mathcal{M}$.

If $\mu_N \rightharpoonup 0$, then $\hat{\mu}_N(\xi) = \mu_N(e^{i2\pi\langle\xi,\cdot\rangle}) \to 0$ for all $\xi \in \mathbb{Z}^d$. Since $|\hat{\mu}_N(\xi)\hat{h}(\xi)| \leq 2M|\hat{h}(\xi)|$ for all $\xi \in \mathbb{Z}^d$ and $\sum_{\xi\in\mathbb{Z}^d} |2M\hat{h}(\xi)|^2 < \infty$, dominated convergence yields that $\mathcal{N}_h(\mu_N) \to 0$.

Conversely, assume that $\mathcal{N}_h(\mu_N) \to 0$. Since the $\mu_N$ are bounded, there are subsequences $\mu_{N_s}$ that converge weakly to a measure $\nu$ that depends on the subsequence. We have to prove that $\nu = 0$ for all such subsequences. Since $\mathcal{N}_h(\mu_N) \to 0$, we have $\hat{\mu}_N(\xi) \to 0$ for all $\xi \in \mathbb{Z}^d$. Therefore, $\hat{\nu}(\xi) = 0$, $\forall \xi \in \mathbb{Z}^d$. This is equivalent to $\nu = 0$ (see e.g. (Katznelson, 1968, p.36)), ending the proof. □

### 5.3.2 Existence of solutions

The first important question one may ask is whether Problem (5.6) admits a solution or not. Theorem 5.6 provides sufficient conditions for existence to hold.

**Proposition 5.6.** *If $\mathcal{M}_N$ is weakly compact, then Problem (5.6) admits at least a solution. In particular, if $\mathcal{M}_N$ is weakly closed and bounded in TV-norm, Problem (5.6) admits at least a solution.*

*Proof.* Assume $\mathcal{M}_N$ is weakly compact. Consider a minimizing sequence $\mu_n \in \mathcal{M}_N$. By compacity, there is a $\mu \in \mathcal{M}_N$ and a subsequence $(\mu_{n_k})_{k\in\mathbb{N}}$ such that $\mu_{n_k} \underset{k\to+\infty}{\rightharpoonup} \mu$. By Proposition 5.5, $\mathcal{N}_h$ induces the weak topology on any TV-bounded set of signed measures, so that $\lim_{k\to\infty} \mathcal{N}_h(\pi - \mu_k) = \mathcal{N}_h(\pi - \mu)$.

Since closed balls in TV-norms are weakly compact, any weakly closed TV-bounded set is weakly compact. □

A key concept that will appear in the continuous line drawing problem is that of push-forward or *empirical measure* (Bogachev and Ruas, 2007) defined hereafter. Let $(X, \gamma)$ denote an arbitrary probability space. Given a function $p : X \to \Omega$, the empirical measure associated with $p$ is denoted $p_*\gamma$. It is defined for any measurable set $B$ by

$$p_*\gamma(B) \coloneqq \gamma(p^{-1}(B)),$$

where $\gamma$ denotes the Lebesgue measure on the interval $[0, 1]$. Intuitively, the quantity $p_*\gamma(B)$ represents the "time" spent by the function $p$ in $B$. Note that $p_*\gamma$ is a probability measure since it is positive and $p_*\gamma(\Omega) = 1$. Given a measure $\mu$ of kind $\mu = p_*\gamma$, the function $p$ is called *parameterization* of $\mu$.

Let $\mathcal{P}$ denote a set of parameterizations $p : X \to \Omega$ and $\mathcal{M}(\mathcal{P})$ denote the associated set of pushforward-measures:

$$\mathcal{M}(\mathcal{P}) \coloneqq \{\mu = p_*\gamma, p \in \mathcal{P}\}.$$

In the rest of this paragraph we give sufficient conditions so that a projection on $\mathcal{M}(\mathcal{P})$ exists. We first need the following proposition.

**Proposition 5.7.** *Let $(p_n)_{n\in\mathbb{N}}$ denote a sequence in $\mathcal{P}$ that converges to $p$ pointwise. Then $(p_{n*}\gamma)_{n\in\mathbb{N}}$ converges weakly to $p_*\gamma$.*

*Proof.* Let $f \in \mathcal{C}(\Omega)$. Since $\Omega$ is compact, $f$ is bounded. Hence dominated convergence yields $\int_X f(p_n(x)) - f(p(x))d\gamma(x) \to 0$. □

**Proposition 5.8.** *Assume that $\mathcal{P}$ is compact for the topology of pointwise convergence. Then there exists a minimizer to Problem (5.6) with $\mathcal{M}_N = \mathcal{M}(\mathcal{P})$.*

*Proof.* By Proposition 5.6 it is enough to show that $\mathcal{M}(\mathcal{P})$ is weakly compact. First, $\mathcal{M}(\mathcal{P})$ is bounded in TV-norm since it is a subspace of probability measures. Consider a

sequence $(p_n)_{n\in\mathbb{N}}$ in $\mathcal{P}$ such that the sequence $(p_{n*}\gamma)_{n\in\mathbb{N}}$ weakly converges to a measure $\mu$. Since $\mathcal{P}$ is compact for the topology of pointwise convergence, there is a subsequence $(p_{n_k})_{k\in\mathbb{N}}$ converging pointwise to $p \in \mathcal{P}$. By Proposition 5.7, the pushforward-measure $p_*\gamma = \mu$ so that $\mu \in \mathcal{M}(\mathcal{P})$ and $\mathcal{P}$ is weakly closed. □

### 5.3.3 Consistency

In this paragraph, we consider a sequence $(\mathcal{M}_N)_{N\in\mathbb{N}}$ of weakly compact subsets of $\mathcal{M}_\Delta$. By Proposition 5.6 there exists a minimizer $\mu_N^* \in \mathcal{M}_N$ to Problem (5.6) for every $N$. We provide a necessary and sufficient condition on $(\mathcal{M}_N)_{N\in\mathbb{N}}$ for consistency, i.e. $\mu_N^* \underset{N\to\infty}{\rightharpoonup} \pi$. In the case of image rendering, it basically means that if $N$ is taken sufficiently large, the projection $\mu_N^*$ and the target image $\pi$ will be indistinguishable from a perceptual point of view. The first result reads as follows.

**Theorem 5.9.** *The following assertions are equivalent:*

*i) For all* $\pi \in \mathcal{M}_\Delta$, $\mu_N^* \underset{N\to\infty}{\rightharpoonup} \pi$.

*ii)* $\cup_{N\in\mathbb{N}}\mathcal{M}_N$ *is weakly dense in* $\mathcal{M}_\Delta$.

*Proof.* We first prove $ii) \Rightarrow i)$. Assume that $\cup_{N\in\mathbb{N}}\mathcal{M}_N$ is weakly dense in $\mathcal{M}_\Delta$. This implies that, $\forall\pi \in \mathcal{M}_\Delta$, $\exists(\mu_N)_{N\in\mathbb{N}} \in (\mathcal{M}_N)_{N\in\mathbb{N}}$ such that $\mu_N \underset{N\to\infty}{\rightharpoonup} \pi$. From Proposition 5.5, this is equivalent to $\lim_{N\to\infty} \mathcal{N}_h(\mu_N - \pi) = 0$. Since $\mu_N^*$ is the projection

$$0 \leq \mathcal{N}_h(\mu_N^* - \pi) \leq \mathcal{N}_h(\mu_N - \pi) \to 0.$$

Proposition 5.5 implies that $\mu_N^* \underset{N\to\infty}{\rightharpoonup} \pi$.

The proof of $i) \Rightarrow ii)$ is straightforward by contraposition. Indeed, if $\cup_{N\in\mathbb{N}}\mathcal{M}_N$ is not weakly dense in $\mathcal{M}_\Delta$, there exists $\pi_0 \in \mathcal{M}_\Delta$ that can not be approximated weakly by any sequence $(\mu_N)_{N\in\mathbb{N}} \in (\mathcal{M}_N)_{N\in\mathbb{N}}$. □

We now turn to the more ambitious goal of assessing the speed of convergence of $\mu_N^*$ to $\pi$. The most natural metric in our context is the minimized norm $\mathcal{N}_h(\mu_N^* - \pi)$. However, its analysis is easy in the Fourier domain, whereas all measures sets in this paper are defined in the space domain. We therefore prefer to use another metrization of weak convergence, given by the transportation distance. Moreover we will see in Theorem 5.11 that the transportation distance defined below dominates $\mathcal{N}_h$.

**Definition 5.10.** The $L^1$ transportation distance, also known as Kantorovitch or Wasserstein distance, between two measures with same TV norm is given by:

$$W_1(\mu,\nu) \coloneqq \inf_c \int \|x-y\|_1 \,\mathrm{d}c(x,y)$$

where the infimum runs over all couplings of $\mu$ and $\nu$, that is the measures $c$ on $\Omega\times\Omega$ with marginals satisfying $c(A,\Omega)=\mu(A)$ and $c(\Omega,A)=\nu(A)$ for all Borelians $A$.

Equivalently, we may define the distance through the dual, that is the action on Lipschitz functions:

$$W_1(\mu,\nu) = \sup_{f:Lip(f)\leq 1} \mu(f)-\nu(f). \tag{5.7}$$

We define the point-to-set distance as

$$W_1(\mathcal{M}_N,\pi) \coloneqq \inf_{\mu\in\mathcal{M}_N} W_1(\mu,\pi).$$

Obviously this distance satisfies:

$$W_1(\mathcal{M}_N,\pi) \leq \delta_N \coloneqq \sup_{\pi\in\mathcal{M}_\Delta} \inf_{\mu\in\mathcal{M}_N} W_1(\mu,\pi). \tag{5.8}$$

**Theorem 5.11.** *Assume that $h\in\mathcal{C}(\Omega)$ denote a Lipschitz continuous function with Lipschitz constant $L$. Then*

$$\mathcal{N}_h(\mu-\pi) \leq LW_1(\mu,\pi) \tag{5.9}$$

*and*

$$\mathcal{N}_h(\mu_N^*-\pi) \leq LW_1(\mathcal{M}_N,\pi) \leq L\delta_N. \tag{5.10}$$

*Proof.* Let $\tau_x : h(\cdot)\mapsto h(x-\cdot)$ denote the symmetrization and shift operator. Let us first prove inequality (5.9):

$$\begin{aligned}
\|h\star(\mu-\pi)\|_2^2 &= \int_\Omega [h\star(\mu-\pi)(x)]^2 \,dx \\
&= \int_\Omega |\mu(\tau_x h)-\pi(\tau_x h)|^2 \,dx \\
&\leq |\Omega| L^2 W_1^2(\mu,\pi),
\end{aligned}$$

where we used the dual definition (5.7) of the Wasserstein distance to obtain the last inequality.

Let $\mu_N$ denote a minimizer of $\inf_{\mu\in\mathcal{M}_N} W_1(\mu,\pi)$. If no minimizer exists we may take an $\epsilon$-solution with arbitrary small $\epsilon$ instead. By definition of the projection $\mu_N^*$, we have:

$$\mathcal{N}_h(\mu_N^* - \pi) \leq \mathcal{N}_h(\mu_N - \pi) \leq W(\mu_N, \pi) \leq \delta_N. \tag{5.11}$$

□

Even though the bound (5.10) is pessimistic in general, it provides some insight on which sequences of measure spaces allow a fast weak convergence.

### 5.3.4 Application to image stippling

In order to illustrate the proposed theory, we first focus on the case of $N$-point measures $\mathcal{M}(\Omega^N)$ defined in Eq. 5.2. This setting is the standard one considered for probability quantization (see (Gruber, 2004; Kloeckner, 2012) for similar results). As mentioned earlier, it has many applications including image stippling. Our main results read as follows.

**Theorem 5.12.** *Let $h$ denote an $L$-Lipschitz kernel. The set of $N$-point measures $\mathcal{M}(\Omega^N)$ satisfies the following inequalities:*

$$\delta_N = \sup_{\pi\in\mathcal{M}_\Delta} \inf_{\mu\in\mathcal{M}(\Omega^N)} W_1(\mu,\pi) \leq \left(\frac{\sqrt{d}}{2}+1\right)\frac{1}{N^{1/d}-1} \tag{5.12}$$

*and*

$$\sup_{\pi\in\mathcal{M}_\Delta} \inf_{\mu\in\mathcal{M}(\Omega^N)} \mathcal{N}_h(\mu-\pi) \leq L\left(\frac{\sqrt{d}}{2}+1\right)\frac{1}{N^{1/d}-1}. \tag{5.13}$$

As a direct consequence, we get the following corollary.

**Corollary 5.13.** *Let $\mathcal{M}_N = \mathcal{M}(\Omega^N)$ denote the set of $N$-point measures. Then there exist solutions $\mu_N^*$ to the projection problem (5.6).*

*Moreover, for any $L$-Lipschitz kernel $h \in \mathcal{C}(\Omega)$:*

*i)* $\mu_N^* \underset{N\to\infty}{\rightharpoonup} \pi$.

*ii)* $\mathcal{N}_h(\mu_N^* - \pi) = \mathcal{O}\left(LN^{-\frac{1}{d}}\right)$.

*Proof.* We first evaluate the bound $\delta_N$ defined in (5.8). To this end, for any given $\pi$, we construct an explicit sequence of measures $\mu_0, \ldots, \mu_N$, the last of which is an $N$-point measure approximating $\pi$.

Note that $\mathbb{T}^d$ can be thought of as the unit cube $[0,1)^d$. It may therefore be partitioned in $C^d$ smaller cubes of edge length $1/C$ with $C = \lfloor N^{1/d} \rfloor$. We let $(\omega_i)_{1 \leqslant i \leqslant C^d}$ denote the small cubes and $x_i$ denote their center. We assume that the cubes are ordered in such a way that $\omega_i$ and $\omega_{i+1}$ are contiguous.

We define $\mu_0 = \sum_{i=1}^{C^d} \pi(\omega_i)\delta_{x_i}$. The measure $\mu_0$ satisfies

$$\begin{aligned} W_1(\pi, \mu_0) &\leqslant \frac{1}{2} \sup_i \text{Diameter}(\omega_i) \\ &\leqslant \frac{\sqrt{d}}{2} \lfloor N^{1/d} \rfloor^{-1} \\ &\leqslant \frac{\sqrt{d}}{2} \frac{1}{N^{1/d} - 1}, \end{aligned}$$

but is not an $N$-point measure since $N\pi(\omega_i)$ is not an integer.

To obtain an $N$-point measure, we recursively build $\mu_l$ as follows:

$$\begin{aligned} \mu_l(\{x_l\}) &= \frac{1}{N} \left\lfloor N\mu_{l-1}(\{x_l\}) \right\rfloor, \\ \mu_l(\{x_{l+1}\}) &= \mu_{l-1}(\{x_{l+1}, x_l\}) - \frac{1}{N} \left\lfloor N\mu_{l-1}(\{x_l\}) \right\rfloor \\ &\qquad\qquad \text{if } l \leq (1/C)^d - 1, \\ \mu_l(\{x_i\}) &= \mu_{l-1}(\{x_i\}) \quad \text{if } i \notin \{l, l+1\}. \end{aligned}$$

We stop the process for $l = (1/C)^d$ and let $\tilde{\mu} = \mu_{(1/C)^d}$. Notice that $N\mu_l(x_i)$ is an integer for all $i \leqslant l$ and that $\mu_l$ is a probability measure for all $l$. Therefore $\tilde{\mu}$ is an $N$-point measure. Moreover:

$$\begin{aligned} W_1(\mu_l, \mu_{l+1}) &\leqslant \frac{1}{N} \|x_l - x_{l+1}\|_2 \\ &\leqslant \frac{1}{N(N^{1/d} - 1)}. \end{aligned}$$

Since the transportation distance is a distance, we have the triangle inequality. Therefore:

$$\begin{aligned} W_1(\pi, \tilde{\mu}) &\leq W_1(\pi, \mu_0) + \sum_{l=1}^{N} W_1(\mu_{l-1}, \mu_l), \\ &= \frac{\sqrt{d}}{2} \frac{1}{N^{1/d} - 1} + N \frac{1}{N(N^{1/d} - 1)} \\ &= \left( \frac{\sqrt{d}}{2} + 1 \right) \frac{1}{N^{1/d} - 1}. \end{aligned}$$

The inequality (5.13) is a direct consequence of this result and Proposition 5.11.

We now turn to the proof of Corollary 5.13. To prove the existence, first notice that the projection problem (5.6) can be recast as (5.1). Let $p = (p_1, \cdots, p_N) \in \Omega^N$. The mapping $p \mapsto \left\| h \star \left( \pi - \frac{1}{N} \sum_{i=1}^N \delta_{p_i} \right) \right\|_2^2$ is continuous. Problem (5.1) therefore consists of minimizing a finite dimensional continuous function over a compact set. The existence of a solution follows. Point ii) is a direct consequence of Theorem 5.11 and bound (5.13). Point i) is due to the fact that $\mathcal{N}_h$ metrizes weak convergence, see Proposition 5.5. □

## 5.4 Numerical resolution

In this section, we propose a generic numerical algorithm to solve the projection problem (5.6). We first draw a connection with the recent works on electrostatic halftoning (Schmaltz et al., 2010; Teuber et al., 2011) in subsection 5.4.1. We establish a connection with Thomson's problem (Thomson, 1904) in subsection 5.4.2. We then recall the algorithm proposed in (Schmaltz et al., 2010; Teuber et al., 2011) when $\mathcal{M}_N$ is the set of $N$-point measures. Finally, we extend this principle to arbitrary measures spaces and provide some results on their theoretical performance in section 5.4.4.

### 5.4.1 Relationship to electrostatic-halftoning

In a recent series of papers (Schmaltz et al., 2010; Teuber et al., 2011; Gräf et al., 2012; Gwosdek et al., 2014), it was suggested to use electrostatic principles to perform image halftoning. This technique was shown to produce results having a number of nice properties such as few visual artifacts and state-of-the-art performance when convolved with a Gaussian filter. Motivated by preliminary results in (Schmaltz et al., 2010), the authors of (Teuber et al., 2011) proposed to choose the $N$ points locations $p = (p_i)_{1 \leq i \leq N} \in \Omega^N$ as a solution of the following variational problem:

$$\min_{p \in \Omega^N} \underbrace{\frac{1}{2N^2} \sum_{i=1}^N \sum_{j=1}^N H(p_i - p_j)}_{\text{Repulsion potential}} - \underbrace{\frac{1}{N} \sum_{i=1}^N \int_\Omega H(x - p_i)\, d\pi(x)}_{\text{Attraction potential}}, \tag{5.14}$$

where $H$ was initially defined as $H(x) = -\|x\|_2$ in (Schmaltz et al., 2010; Teuber et al., 2011) and then extended to a few other functions in (Gräf et al., 2012). The attraction potential tends to attract points towards the bright regions of the image (regions where the measure $\pi$ has a large mass) whereas the repulsion potential can be regarded as a counter-balancing term that tends to maximize the distance between all pairs of points.

Proposition 5.14 below shows that this attraction-repulsion problem is actually equivalent to the projection problem (5.6) on the set of $N$-point measures defined in (5.2). We let

$\mathcal{P}^*$ denote the set of solutions of (5.14) and $\mathcal{M}(\mathcal{P}^*) = \{\mu = \frac{1}{N}\sum_{i=1}^N \delta_{p_i^*},\ p^* \in \mathcal{P}^*\}$. We also let $\mathcal{M}^*$ denote the set of solutions to problem (5.6).

**Proposition 5.14.** *Let $h \in \mathcal{C}(\Omega)$ denote a kernel such that $|\hat{h}|(\xi) > 0,\ \forall \xi \in \mathbb{Z}^d$. Define $H$ through its Fourier series by $\hat{H}(\xi) = |\hat{h}|^2(\xi)$. Then problems (5.6) and (5.14) yield the same solutions set:*

$$\mathcal{M}^* = \mathcal{M}(\mathcal{P}^*).$$

*Proof.* First, note that since $H$ and $h$ are continuous both problems are well defined and admit at least one solution. Let us first expand the $L^2$-norm in (5.6):

$$\begin{aligned}\frac{1}{2}\|h \star (\mu - \pi)\|_2^2 &= \frac{1}{2}\langle h \star (\mu - \pi), h \star (\mu - \pi)\rangle \\ &= \frac{1}{2}\langle H \star (\mu - \pi), \mu - \pi\rangle \\ &= \frac{1}{2}\left(\langle H \star \mu, \mu\rangle - 2\langle H \star \mu, \pi\rangle + \langle H \star \pi, \pi\rangle\right).\end{aligned}$$

Therefore

$$\operatorname*{Arg\,min}_{\mu \in \mathcal{M}_N} \frac{1}{2}\|h \star (\mu - \pi)\|_2^2 = \operatorname*{Arg\,min}_{\mu \in \mathcal{M}_N} \frac{1}{2}\left(\langle H \star \mu, \mu\rangle - 2\langle H \star \mu, \pi\rangle\right).$$

To conclude, it suffices to remark that for a measure $\mu$ of kind $\mu = \frac{1}{N}\sum_{i=1}^N \delta_{p_i}$,

$$\begin{aligned}&\frac{1}{2}\left(\langle H \star \mu, \mu\rangle - 2\langle H \star \mu, \pi\rangle\right) \\ &= \frac{1}{2N^2}\sum_{i=1}^N\sum_{j=1}^N H(p_i - p_j) - \frac{1}{N}\sum_{i=1}^N \int_\Omega H(x - p_i)\, d\pi(x).\end{aligned}$$

□

*Remark* 5.15. It is rather easy to show that a sufficient condition for $h$ to be continuous is that $H \in \mathcal{C}^3(\Omega)$ or $H$ be Hölder continuous with exponent $\alpha > 2$. These conditions are however strong and exclude kernels such as $H(x) = -\|x\|_2$.

From Remark 5.2, it is actually sufficient that $h \in \mathcal{L}^\infty(\Omega)$ for $\mathcal{N}_h$ to be well defined. This leads to less stringent conditions on $H$. We do not discuss this possibility further to keep the arguments simple.

*Remark* 5.16. Corollary 5.13 sheds light on the approximation quality of the minimizers of attraction-repulsion functionals. Let us mention that consistency of problem (5.14) was already studied in the recent papers (Gräf et al., 2012; Fornasier et al., 2013; Fornasier and Hütter, 2013). To the best of our knowledge, Corollary 5.13 is stronger than existing results since it yields a convergence rate and holds true under more general assumptions.

Though formulations (5.6) and (5.14) are equivalent, we believe that the proposed one (5.6) has some advantages: it is probably more intuitive, shows that the convolution kernel $h$ should be chosen depending on physical considerations and simplifies some parts of the mathematical analysis such as consistency. However, the set of admissible measures $\mathcal{M}(\Omega^N)$ has a complex geometry and this formulation as such is hardly amenable to numerical implementation. For instance, $\mathcal{M}(\Omega^N)$ is not a vector space, since adding two $N$-point measures usually leads to $(2N)$-point measures. On the other hand, the attraction-repulsion formulation (5.14) is an optimization problem of a continuous function over the set $\Omega^N$. It therefore looks easier to handle numerically using non-linear programming techniques. This is what we will implement in the next paragraphs following previous works (Schmaltz et al., 2010; Teuber et al., 2011).

### 5.4.2 Link with Thomson's problem

Before going further into the design of a numerical algorithm, let us first show that a specific instance of problem (5.6) is equivalent to Thomson's problem (Thomson, 1904). This is a longstanding open problem in numerical optimization. It belongs to Smale's list of mathematical questions to solve for the XXIst century (Smale, 1998). A detailed presentation of Thomson's problem and its extensions is also proposed in (Hiriart-Urruty, 2009).

Let $\mathbb{S} = \{p \in \mathbb{R}^3, \|p\|_2 = 1\}$ denote the unit 3-dimensional sphere. Thomson's problem may be enounced as follows:

$$\text{Find} \quad p \in \underset{(p_1,\dots,p_N)\in\mathbb{S}^N}{\operatorname{Arg\,min}} \sum_{i\neq j} \frac{1}{\|p_i - p_j\|_2}. \tag{5.15}$$

The term $\sum_{i\neq j} \frac{1}{\|p_i-p_j\|_2}$ represents the electrostatic potential energy of $N$ electrons. Thomson's problem therefore consists of finding the minimum energy configuration of $N$ electrons on the sphere $\mathbb{S}$.

To establish the connection between (5.6) and (5.15), it suffices to set $H(x) = \frac{1}{\|x\|_2}$, $\Omega = \mathbb{S}$ and $\pi = 1$ in Eq. (5.14). By doing so, the attraction potential has the same value whatever the points configuration and the repulsion potential exactly corresponds to the electrostatic potential.

This simple remark shows that finding global minimizers looks too ambitious in general and we will therefore concentrate on the search of local minimizers only.

### 5.4.3 The case of $N$-point measures

In this section, we develop an algorithm specific to the projection on the set of $N$-point measures defined in (5.2). This algorithm generates stippling results such as in Fig. 5.1. In stippling, the measure is supported by a union of discs, i.e., a sum of diracs convoluted with a disc indicator. We simply have to consider the image deconvoluted with this disc indicator as $\pi$ to include stippling in the framework of $N$-point measures. We will generalize this algorithm to arbitrary sets of measures in the next section. We assume without further mention that $\hat{H}(\xi)$ is real and positive for all $\xi$. This implies that $H$ is real and even. Moreover, Proposition 5.14 implies that problems (5.6) and (5.14) yield the same solutions sets. We let $p = (p_1, \dots, p_N)$ and set

$$\tilde{J}(p) \coloneqq \underbrace{\frac{1}{2N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} H(p_i - p_j)}_{F(p)} - \underbrace{\frac{1}{N} \sum_{i=1}^{N} \int_{\Omega} H(x - p_i)\, d\pi(x)}_{\tilde{G}(p)}. \tag{5.16}$$

The projection problem therefore rereads as:

$$\min_{p \in \Omega^N} \tilde{J}(p). \tag{5.17}$$

For practical purposes, the integrals in $\tilde{G}(p)$ first have to be replaced by numerical quadratures. We let $G(p) \simeq \tilde{G}(p)$ denote the numerical approximation of $\tilde{G}(p)$. This approximation can be written as

$$G(p) = \frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{n} w_j H(x_j - p_i) \pi_j,$$

where $n$ is the number of discretization points $x_j$ and $w_j$ are weights that depend on the integration rule. In particular, since we want to approximate integration with respect to a probability measure, we require that

$$\sum_{j=1}^{n} w_j \pi_j = 1.$$

In our numerical experiments we use the rectangle rule. We may then take $\pi_j$ as the integral of $\pi$ over the corresponding rectangle. After discretization, the projection problem therefore rereads as:

$$\min_{p \in \Omega^N} J(p) \coloneqq F(p) - G(p). \tag{5.18}$$

The following result (Attouch et al., 2013, Theorem 5.3) will be useful to design a convergent algorithm. We refer to (Attouch et al., 2013) for a comprehensive introduction

to the definition of Kurdyka-Łojasiewicz functions and to its applications to algorithmic analysis. In particular, we recall that semi-algebraic functions are Kurdyka-Łojasiewicz (Kurdyka, 1998).

**Theorem 5.17.** *Let $K : \mathbb{R}^n \to \mathbb{R}$ be $\mathcal{C}^1$ function whose gradient is L-Lipschitz continuous and let $C$ be a nonempty closed subset of $\mathbb{R}^n$. Being given $\varepsilon \in \left(0, \frac{1}{2L}\right)$ and a sequence of stepsizes $\gamma^{(k)}$ such that $\varepsilon < \gamma^{(k)} < \frac{1}{L} - \varepsilon$, we consider a sequence $(x^{(k)})_{k\in\mathbb{N}}$ that complies with*

$$x^{(k+1)} \in P_C\left(x^{(k)} - \gamma^{(k)}\nabla K(x^{(k)})\right), \text{ with } x^{(0)} \in C \tag{5.19}$$

*If the function $K + i_C$ is a Kurdyka-Łojasiewicz function and if $(x^{(k)})_{k\in\mathbb{N}}$ is bounded, then the sequence $(x^{(k)})_{k\in\mathbb{N}}$ converges to a critical point $x^*$ in $C$.*

A consequence of this important result is the following.

**Corollary 5.18.** *Assume that $H$ is a $\mathcal{C}^1$ semi-algebraic function with L-Lipschitz continuous gradient. Set $0 < \gamma < \frac{N}{3L}$. Then the following sequence converges to a critical point of problem (5.18)*

$$p^{(k+1)} \in P_{\Omega^N}\left(p^{(k)} - \gamma\nabla J(p^{(k)})\right), \text{ with } p^{(0)} \in \Omega^N. \tag{5.20}$$

*If $H$ is convex, $0 < \gamma < \frac{N}{2L}$ ensures convergence to a critical point.*

*Remark* 5.19. The semi-algebraicity is useful to obtain convergence to a critical point. In some cases it might however not be needed. For instance, in the case where $C$ is convex and closed, it is straightforward to establish the decrease of the cost function assuming only that $\nabla J$ is Lipschitz. Nesterov in (Nesterov, 2013, Theorem 3) also provides a convergence rate in $\mathcal{O}\left(\frac{1}{\sqrt{k+1}}\right)$ in terms of objective function values.

*Proof.* First notice that $J$ is semi-algebraic as a finite sum of semi-algebraic functions.

Function $J$ is $\mathcal{C}^1$ by Leibniz integral rule. Let $\partial_k$ denote the derivative with respect to $p_k$. Then, since $H$ is even

$$\partial_k F(p) = \frac{1}{N^2}\sum_{i=1}^{N}\nabla H(p_k - p_i) \tag{5.21}$$

and

$$\partial_k G(p) = -\frac{1}{N}\sum_{j=1}^{n} w_j \nabla H(x_j - p_k)\pi_j. \tag{5.22}$$

For any two sets of $N$ points $p^{(1)} = (p_k^{(1)})_{1\leqslant k\leqslant N},\ p^{(2)} = (p_k^{(2)})_{1\leqslant k\leqslant N}$:

$$\begin{aligned}
\|\nabla F(p^{(1)}) - \nabla F(p^{(2)})\|_2^2 &= \sum_{k=1}^N \Big\|\partial_k F(p^{(1)}) - \partial_k F(p^{(2)})\Big\|_2^2 \\
&= \frac{1}{N^4}\sum_{k=1}^N \Big\| \sum_{i=1}^N \nabla H(p_k^{(1)} - p_i^{(1)}) - \nabla H(p_k^{(2)} - p_i^{(2)})\Big\|_2^2 \\
&\leqslant \frac{1}{N^4}\sum_{k=1}^N \Big( \sum_{i=1}^N L\|p_k^{(1)} - p_i^{(1)} - (p_k^{(2)} - p_i^{(2)})\|_2\Big)^2 \\
&\leqslant \frac{L^2}{N^4}\sum_{k=1}^N \Big( \sum_{i=1}^N \|p_k^{(1)} - p_k^{(2)}\|_2 + \|p_i^{(1)} - p_i^{(2)}\|_2\Big)^2 \\
&\leqslant \frac{L^2}{N^4}\sum_{k=1}^N N\Big( \sum_{i=1}^N \big(\|p_k^{(1)} - p_k^{(2)}\|_2 + \|p_i^{(1)} - p_i^{(2)}\|_2\big)^2\Big) \\
&\leqslant \frac{2L^2}{N^3}\sum_{k=1}^N\sum_{i=1}^N \|p_k^{(1)} - p_k^{(2)}\|_2^2 + \|p_i^{(1)} - p_i^{(2)}\|_2^2 \\
&= \frac{4L^2}{N^2}\|p^{(1)} - p^{(2)}\|_2^2,
\end{aligned}$$

and

$$\begin{aligned}
\|\nabla G(p^{(1)}) - \nabla G(p^{(2)})\|_2^2 &= \sum_{k=1}^N \Big\|\partial_k G(p^{(1)}) - \partial_k G(p^{(2)})\Big\|_2^2 \\
&= \frac{1}{N^2}\sum_{k=1}^N \Big\| \sum_{j=1}^n w_j\pi_j\big(\nabla H(p_k^{(1)} - x) - \nabla H(p_k^{(2)} - x)\big)\Big\|_2^2 \\
&\leqslant \frac{1}{N^2}\sum_{k=1}^N \Big( \sum_{j=1}^n w_j\pi_j L\|p_k^{(1)} - p_k^{(2)}\|\Big)^2 \\
&= \frac{L^2}{N^2}\Big(\sum_{j=1}^n w_j\pi_j\Big)\|p^{(1)} - p^{(2)}\|_2^2 \\
&= \frac{L^2}{N^2}\|p^{(1)} - p^{(2)}\|_2^2.
\end{aligned}$$

Finally,

$$\begin{aligned}
&\|\nabla J(p^{(1)}) - \nabla J(p^{(2)})\|_2 \\
&\leqslant \|\nabla F(p^{(1)}) - \nabla F(p^{(2)})\|_2 + \|\nabla G(p^{(1)}) - \nabla G(p^{(2)})\|_2 \\
&\leqslant \Big(\frac{2L}{N} + \frac{L}{N}\Big)\|p^{(1)} - p^{(2)}\|_2 = \frac{3L}{N}\|p^{(1)} - p^{(2)}\|_2.
\end{aligned}$$

Now, if we assume that $H$ is convex and $\mathcal{C}^2$ (this hypothesis is not necessary, but simplifies the proof). Then $F$ and $G$ are also convex and $\mathcal{C}^2$. We let $\nabla^2 F$ denote the Hessian matrix

of $F$. Given the previous inequalities, we have $0 \preccurlyeq \nabla^2 F \preccurlyeq \frac{2L}{N}\mathrm{Id}$ and $0 \preccurlyeq \nabla^2 G \preccurlyeq \frac{L}{N}\mathrm{Id}$. Hence, the largest eigenvalue in magnitude of $\nabla^2(F-G)$ is bounded above by $\frac{2L}{N}$.

Moreover, the sequence $(x^{(k)})_{k\in\mathbb{N}}$ is bounded since $\Omega^N$ is bounded. □

### 5.4.4 A generic projection algorithm

We now turn to the problem of finding a solution of (5.6), where $\mathcal{M}_N$ denotes our arbitrary measures set. In the previous paragraph, it was shown that critical points of $J + i_{\Omega^N}$ could be obtained with a simple projected gradient algorithm under mild assumtpions. Although this algorithm only yields critical points, they usually correspond to point configurations that are visually pleasing after only a few hundreds of iterations. For instance, the lion in Figure 5.1b was obtained after 200 iterations. Motivated by this appealing numerical behavior, we propose to extend this algorithm to the following abstract construction:

1. Approximate $\mathcal{M}_N$ by a subset $\mathcal{A}_n$ of $n$-point measures.
2. Use the generic Algorithm (5.19) to obtain an approximate projection $\mu_n^*$ on $\mathcal{A}_n$.
3. When possible, reconstruct an approximation $\mu_N \in \mathcal{M}_N$ of a projection $\mu_N^*$ using $\mu_n^*$.

To formalize the approximation step, we need the definition of Hausdorff distance:

**Definition 5.20.** The Hausdorff distance between two subsets $X$ and $Y$ of a metric space $(M, d)$ is:

$$\mathcal{H}_d(X,Y) := \max\left\{\sup_{x\in X}\inf_{y\in Y} d(x,y), \sup_{y\in Y}\inf_{x\in X} d(y,x)\right\}.$$

In words, two sets are close if any point in one set is close to at least a point in the other set. In this paper, the relevant metric space is the space of signed measures $\mathcal{M}$ with the norm $\mathcal{N}_h$. The corresponding Hausdorff distance is denoted $\mathcal{H}_{\mathcal{N}_h}$.

The following proposition clarifies why controlling the Hausdorff distance is relevant to design approximation sets $\mathcal{A}_n$.

**Proposition 5.21.** *Let $\mathcal{A}_n$ and $\mathcal{M}_N$ be two TV-bounded weakly closed sets of measures such that $\mathcal{H}_{\mathcal{N}_h}(\mathcal{A}_n, \mathcal{M}_N) \leq \varepsilon$. Let $\mu_n^*$ be a projection on $\mathcal{A}_n$. Then there is a point $\mu_N \in \mathcal{M}_N$ such that $\mathcal{N}_h(\mu_n^* - \mu_N) \leq \varepsilon$ and $\mathcal{N}_h(\pi - \mu_N) \leq \inf_{\mu\in\mathcal{M}_N} \mathcal{N}_h(\pi - \mu) + 2\varepsilon$.*

**Corollary 5.22.** *If* $\lim_{n\to\infty} \mathcal{H}_{\mathcal{N}_h}(\mathcal{A}_n, \mathcal{M}_N) = 0$, *then* $(\mu_n^*)_{n\in\mathbb{N}}$ *converges weakly along a subsequence to a solution* $\mu_N^*$ *of Problem* (5.6).

*Proof.* We first prove Proposition 5.21. Since $\mathcal{A}_n$ and $\mathcal{M}_N$ are bounded weakly closed, by Proposition 5.6, there exists at least one projection $\mu_n^*$ on $\mathcal{A}_n$ and one projection $\mu_N^*$ on $\mathcal{M}_N$.

Moreover since $\mathcal{A}_n$ and $\mathcal{M}_N$ are bounded weakly closed, they are also closed for $\mathcal{N}_h$, so that the infimum in the Hausdorff distances are attained. Hence there exists $\mu_n \in \mathcal{A}_n$ such that $\mathcal{N}_h(\mu_n - \mu_N^*) \leq \mathcal{H}_{\mathcal{N}_h}(\mathcal{A}_n, \mathcal{M}_N) \leq \varepsilon$ and $\mu_N \in \mathcal{M}_N$ such that $\mathcal{N}_h(\mu_N - \mu_n^*) \leq \varepsilon$. The proposition follows from the triangle inequality:

$$\begin{aligned}
\mathcal{N}_h(\mu_N - \pi) &\leq \mathcal{N}_h(\mu_N - \mu_n^*) + \mathcal{N}_h(\mu_n^* - \pi) \\
&\leq \varepsilon + \mathcal{N}_h(\mu_n - \pi) \\
&\leq \varepsilon + \mathcal{N}_h(\mu_n - \mu_N^*) + \mathcal{N}_h(\mu_N^* - \pi) \\
&\leq \mathcal{N}_h(\mu_N^* - \pi) + 2\varepsilon.
\end{aligned}$$

For the corollary, let us consider the sequence $(\mu_n^*)_{n\in\mathbb{N}}$ as $n$ tends to infinity. Since all $\mu_n$ are in $\mathcal{M}_\Delta$, which is weakly compact, we have a subsequence that converges to $\mu_\infty^*$. Since $\mathcal{N}_h$ is a metrization of weak convergence on $\mathcal{M}_N$, this $\mu_\infty^*$ is indeed a solution to Problem (5.6):

$$\begin{aligned}
\mathcal{N}_h(\mu_\infty^* - \pi) &= \lim_{n\to\infty} \mathcal{N}_h(\mu_n^* - \pi) \\
&= \inf_{\mu\in\mathcal{M}_N} \mathcal{N}_h(\pi - \mu).
\end{aligned}$$

□

To conclude this section, we show that it is always possible to construct an approximation set $\mathcal{A}_n \subseteq \mathcal{M}(\Omega^n)$ with a control on the Hausdorff distance to $\mathcal{M}_N$. Let $\mathcal{M}_N^\epsilon$ denote an $\epsilon$-enlargement of $\mathcal{M}_N$ w.r.t. the $\mathcal{N}_h$-norm, i.e.:

$$\mathcal{M}_N^\epsilon = \cup_{\mu_N\in\mathcal{M}_N} \{\mu \in \mathcal{M}_\Delta, \mathcal{N}_h(\mu - \mu_N) \leq \epsilon\}. \tag{5.23}$$

We may define an approximation set $\mathcal{A}_n^\epsilon$ as follows:

$$\mathcal{A}_n^\epsilon = \mathcal{M}(\Omega^n) \cap \mathcal{M}_N^\epsilon. \tag{5.24}$$

For sufficient large $n$, this set is non-empty and can be rewritten as

$$\mathcal{A}_n^\epsilon = \left\{ \mu = \frac{1}{n} \sum_{i=1}^{n} \delta_{p_i}, \quad \text{with } p = (p_i)_{1 \leq i \leq n} \in \mathcal{P}_n^\epsilon \right\}, \tag{5.25}$$

where the parameterization set $\mathcal{P}_n^\epsilon$ depends on $\mathcal{M}_N$ and $\epsilon$. With this discretization of $\mathcal{M}_N$ at hand, one can then apply (at least formally) the following projected gradient descent algorithm:

$$p^{(k+1)} \in P_{\mathcal{P}_n^\epsilon} \left( p^{(k)} - \gamma \nabla J(p^{(k)}) \right), \text{ with } p^{(0)} \in \mathcal{P}_n^\epsilon. \tag{5.26}$$

The following proposition summarizes the main approximation result:

**Proposition 5.23.** *Assume that $h$ is $L$-Lipschitz. Set $\epsilon = \left( \frac{\sqrt{d}}{2} + 1 \right) \frac{L}{n^{1/d} - 1}$ and $\mathcal{A}_n = \mathcal{A}_n^\epsilon$, then*

$$\mathcal{H}_{\mathcal{N}_h} (\mathcal{A}_n, \mathcal{M}_N) = \mathcal{O} \left( L n^{-1/d} \right).$$

*Proof.* By construction, $\mathcal{A}_n$ satisfies

$$\sup_{\mu_n \in \mathcal{A}_n} \inf_{\mu_N \in \mathcal{M}_N} \mathcal{N}_h(\mu_n - \mu_N) \leq \epsilon.$$

Let $\mu_N$ be an arbitrary measure in $\mathcal{M}_N$. By inequality (5.12), there exists $\mu_n \in \mathcal{M}(\Omega^n)$ such that $\mathcal{N}_h(\mu_n - \mu_N) \leq \epsilon$. Therefore $\mu_n$ also belongs to $\mathcal{A}_n^\epsilon$. This shows that

$$\sup_{\mu_N \in \mathcal{M}_N} \inf_{\mu_n \in \mathcal{A}_n} \mathcal{N}_h(\mu_n - \mu_N) \leq \epsilon.$$

□

The approximation process proposed (5.24) is non-constructive in the does not induce any explicit formula for $\mathcal{P}_n^\epsilon$. Moreover, $\mathcal{P}_n^\epsilon$ can be an arbitrary set and the projection on $\mathcal{P}_n^\epsilon$ might not be implementable. We will provide constructive approximations for specific measures spaces in Section 5.5.

## 5.5 Application to continuous line drawing

In this section, we concentrate on the continuous line drawing problem described in the introduction. We first construct a set of admissible measures $\mathcal{M}_T$ that is a natural representative of artistic continuous line drawings. The index $T$ represents the time spent to draw the picture. We then show that using this set in problem (5.6) ensures

existence of a solution and weak convergence of the minimizers $\mu_T^*$ to any $\pi \in \mathcal{M}_\Delta$. We finish by designing a numerical algorithm to solve the problem and analyze its theoretical guarantees.

### 5.5.1 Problem formalization

Let us assume that an artist draws a picture with a pencil. The trajectory of the pencil tip can be defined as a parameterized curve $p : [0,T] \to \Omega$. The body, elbow, arm and hand are subject to non-trivial constraints (Marteniuk et al., 1987). The curve $p$ should therefore belong to some admissible parameterized curves set denoted $\mathcal{P}_T$. In this paper, we simply assume that $\mathcal{P}_T$ contains curves with bounded first and second order derivatives in $L^q([0,T])$. More precisely, we consider the following sets of admissible curves:

1. Curves with bounded speed:

$$\mathcal{P}_T^{1,\infty} = \Big\{ p \in (W^{1,\infty}([0,T]))^d,\ p([0,T]) \subset \Omega, \|\dot{p}\|_\infty \leq \alpha_1 \Big\},$$

   where $\alpha_1$ is a positive real.

2. Curves with bounded first and second-order derivatives:

$$\mathcal{P}_T^{2,\infty} = \Big\{ p \in (W^{2,\infty}([0,T]))^d,\ p([0,T]) \subset \Omega, \|\dot{p}\|_\infty \leq \alpha_1, \\ \|\ddot{p}\|_\infty \leq \alpha_2 \Big\},$$

   where $\alpha_1$ and $\alpha_2$ are positive reals. This set models rather accurately kinematic constraints that are met in vehicles. It is obviously a rough approximation of arm constraints.

3. The proposed theory and algorithm apply to a more general setting. For instance they cover the case of curves with derivatives up to an arbitrary order bounded in $L^q$ with $q \in [1,\infty]$. We let

$$\mathcal{P}_T^{m,q} = \Big\{ p \in (W^{m,q}([0,T]))^d,\ p([0,T]) \subset \Omega, \\ \forall i \in \{1,\dots,m\},\ \|p^{(i)}\|_q \leq \alpha_i \Big\}.$$

   where $(\alpha_i)_{i=1\dots m}$ are positive reals. This case will be treated only in the numerical experiments to illustrate the variety of results that can be obtained in applications.

Note that all above mentionned sets are convex. The convexity property will help deriving efficient numerical procedures.

In the rest of this section, we consider the following projection problem:

$$\inf_{\mu \in \mathcal{M}\left(\mathcal{P}_T^{m,q}\right)} \mathcal{N}_h(\mu - \pi), \tag{5.27}$$

with a special emphasis on the set $\mathcal{M}\left(\mathcal{P}_T^{m,\infty}\right)$ since it best describes standard kinematic constraints. This problem basically consists of finding the "best" way to represent a picture in a given amount of time $T$.

### 5.5.2 Existence and consistency

We first provide existence results using the results derived in Section 5.3 for $q = \infty$.

**Theorem 5.24.** *For any $m \in \mathbb{N}^*$, Problem (5.27) admits at least one solution in $\mathcal{M}\left(\mathcal{P}_T^{m,\infty}\right)$.*

*Proof.* From Proposition 5.8, it suffices to show that $\mathcal{P}_T^{m,\infty}$ is compact for the topology of pointwise convergence.

Let $(p_n)_{n\in\mathbb{N}}$ be a sequence in $\mathcal{P}_T^{m,\infty}$ that converges pointwise to $p$. Since $p_n$ is in $W^{m,\infty}$, its $(m-1)$-th derivative is Lipschitz continuous. By definition of $\mathcal{P}_T^{m,\infty}$, the $p_n^{(m-1)}$ are both uniformly bounded by $\alpha_{m-1}$ and $\alpha_m$-Lipschitz, hence equicontinuous. Next, by Ascoli's theorem, up to taking a subsequence, $p_n^{(m-1)}$ uniformly converges to a continuous $p^{(m-1)}$. Integrating yields that $p_n^{(i)} \to p^{(i)}$ uniformly for all $i \leq m-1$, so that $\left\|p^{(i)}\right\|_\infty \leq \alpha_i$ for $i \leq m-1$. Finally, a limit of $L$-Lipschitz functions is also $L$-Lipschitz, so that $\left\|p^{(m)}\right\|_\infty \leq \alpha_m$. Hence $p \in \mathcal{P}_T^{m,\infty}$, ending the proof. □

Let us now turn to weak convergence.

**Theorem 5.25.** *Let $T$ be an arbitrary positive real. Let $\mu_T^* \in \mathcal{M}\left(\mathcal{P}_T^{m,\infty}\right)$ denote any solution of Problem (5.27). Then, for any Lipschitz kernel $h \in \mathcal{C}(\Omega)$:*

*i)* $\mu_T^* \underset{T\to\infty}{\rightharpoonup} \pi$,

*ii)* $\mathcal{N}_h(\mu_T^* - \pi) = \mathcal{O}\left(T^{-\frac{m}{m(d+1)-1}}\right)$.

*Proof.* Let us consider a function $u : [0,1] \to \mathbb{R}$ such that:

- The $m$-th derivative is bounded by $\alpha_m$, that is $\left\|u^{(m)}\right\|_\infty \leq \alpha_m$.
- For all integers $i \in \{1, \ldots, m-1\}$, endpoint values are zero, that is $u^{(i)}(0) = u^{(i)}(1) = 0$.

- Start point is zero, that is $u(0) = 0$.
- Endpoint is positive, that is $u(1) = C > 0$.

Let $x$ and $y$ in $\Omega$, such that $\|x - y\|_2 = Cr^m$, and let $\tau_{xy}$ be the unit vector from $x$ to $y$. Then, for $r$ small enough, the function $s[x, y] : t \mapsto x + \tau_{xy} u(\frac{t}{r})$ belongs to $\mathcal{P}_T^{m,\infty}$, with all its first $(m-1)$ derivatives zero at its endpoints. The condition $r$ small enough is for controlling the norm of the $i$-th derivatives for $i \leq m-1$, which scale as $r^{m-i}$.

Now, let us split $\Omega = [0,1]^d$ in $N^d$ small cubes $\omega_i$. We may order them such that each $\omega_i$ is adjacent to the next cube $\omega_{i+1}$. We write $x_i$ for the center of $\omega_i$. We now build functions $s \in \mathcal{P}_T^{m,\infty}$ by concatenating paths from $x_i$ to $x_{i+1}$ and waiting times in $x_i$:

$$
\begin{aligned}
&0 = t_1^1 \leq \cdots \leq t_{i-1}^2 \leq t_i^1 \leq t_i^2 \leq t_{i+1}^1 \leq \cdots \leq t_{N^d}^2 = T, \\
&t_i^2 - t_i^1 = \left(\frac{1}{NC}\right)^{\frac{1}{m}}, \\
&s(t) = \begin{cases} x_i & \text{if} \quad t_i^1 \leq t \leq t_i^2, \\ s[x_i, x_{i+1}](t - t_i^2) & \text{if} \quad t_i^2 \leq t \leq t_{i+1}^1, \end{cases}
\end{aligned}
$$

under the condition $T \geq T_N \coloneqq (N^d - 1)\left(\frac{1}{NC}\right)^{\frac{1}{m}}$, that is to say that we have enough time to loop through all the cube centers.

Let now $\pi \in \mathcal{M}_\Delta$. We may choose $t_i^2 - t_i^1 \leq T\pi(\omega_i)$ for all $i$. Then, we may couple $\pi$ and $s_* \gamma_T$ with $c(x_i, \omega_i) = \frac{t_i^2 - t_i^1}{T}$. Since the small cubes have radius $\sqrt{d}/N$ and the big one has radius $\sqrt{d}$, we obtain:

$$
\begin{aligned}
W_1(\pi, s_*\gamma_T) &\leq \frac{\sqrt{d}}{2N} \sum_i \frac{t_i^2 - t_i^1}{T} + \sqrt{d} \sum_{i < N^d} \frac{t_{i+1}^1 - t_i^2}{T} \\
&= \frac{\sqrt{d}}{2N} \frac{T - T_N}{T} + \sqrt{d} \frac{T_N}{T}.
\end{aligned}
$$

In particular, taking $N = T^{\frac{m}{m(d+1)-1}}$, we find that $W_1\left(\mathcal{M}\left(\mathcal{P}_T^{m,\infty}\right), \pi\right) = O\left(T^{-\frac{m}{m(d+1)-1}}\right)$, hence $\bigcup_T \mathcal{M}\left(\mathcal{P}_T^{m,\infty}\right)$ is weakly dense in $\mathcal{M}_\Delta$.

□

### 5.5.3 Numerical resolution

We now turn to the numerical resolution of problem (5.27). We first discretize the problem. We set $\Delta t := \frac{T}{N}$ and define discrete curves $s$ as vectors of $\mathbb{R}^{N \cdot d}$. We let $s(i) \in \mathbb{R}^d$ denote the curve location at discrete time $i$, corresponding to the continuous time $i\Delta t$.

We define $D_1 : \mathbb{R}^{N\cdot d} \to \mathbb{R}^{N\cdot d}$, the discrete first order derivative operator, as follows:

$$(D_1 s)(i) = \frac{1}{\Delta t} \begin{cases} 0 & \text{if } i = 1, \\ s(i) - s(i-1) & \text{if } i \in \{2, \dots, N\}. \end{cases}$$

In what follows, $D_i$ denotes a discretization of the derivative operator of order $i$. In the numerical experiments, we set $D_2 = -D_1^* D_1$.

We define $P_N^{m,q}$, a discretized version of $\mathcal{P}_T^{m,q}$, as follows:

$$P_N^{m,q} = \big\{ s \in \mathbb{R}^{N\cdot d}, \text{such that } \forall i \in \{1, \dots N\},\ s(i) \in \Omega, \tag{5.28}$$

$$\text{and } \forall j \in \{1, \dots, m\},\ \|D_j s\|_q \leqslant \alpha_j \big\}. \tag{5.29}$$

Here, $\|\cdot\|_q$ is defined by: $\|x\|_q = \left( \sum_{i=1}^{N\cdot d} \|x_i\|_2^q \right)^{\frac{1}{q}}$ for $q \in [1, +\infty)$ and $\|x\|_\infty = \max_{1 \leqslant i \leqslant N\cdot d} \|x_i\|_2$.

The measures set $\mathcal{M}(\mathcal{P}_T^{m,q})$ can be approximated by the set of $N$-point measures $\mathcal{M}(P_N^{m,q})$. From Corollary 5.22, it suffices to control the Hausdorff distance $\mathcal{H}_{W_1}(\mathcal{M}(\mathcal{P}_T^{m,q}), \mathcal{M}(P_N^{m,q}))$, to ensure that the solution of the discrete problem (5.6) with $\mathcal{M}_N = \mathcal{M}(P_N^{m,q})$ is a good approximation of problem (5.27). Unfortunately, the control of this distance is rather technical and falls beyond the scope of this paper for general $m$ and $q$. In the following proposition, we therefore limit ourselves to the case $m = 1, q = \infty$.

**Proposition 5.26.** $\mathcal{H}_{W_1}(\mathcal{M}(\mathcal{P}_T^{1,\infty}), \mathcal{M}(P_N^{1,\infty})) \leqslant \alpha_1 \frac{T}{N}$.

*Proof.* 1. Let us show that $\sup_{\mu \in \mathcal{M}(\mathcal{P}_T^{1,\infty})} \inf_{\tilde{\mu} \in \mathcal{M}(P_N^{1,\infty})} W_1(\mu, \tilde{\mu}) \leqslant \frac{\alpha_1 T}{N}$.

Let $\mu \in \mathcal{M}(\mathcal{P}_T^{1,\infty})$ and denote by $p \in \mathcal{P}_T^{1,\infty}$ a parameterization such that $\mu = p_* \gamma$. Define $\tilde{\mu} = \frac{1}{N} \sum_{i=0}^{N-1} \delta_{p(\frac{iT}{N})}$. Then a parameterization of $\tilde{\mu}$ is defined by $s(i) = p\left(\frac{iT}{N}\right)$. Moreover, for $i \in \{2, \dots N\}$, $|(D_1 s)(i)| = \frac{1}{\Delta t} \left| p\left(\frac{iT}{N}\right) - p\left(\frac{(i-1)T}{N}\right) \right| = \frac{1}{\Delta t} \left| \int_{\frac{(i-1)T}{N}}^{\frac{iT}{N}} \dot{p}(t)\, dt \right| \leqslant \frac{1}{\Delta t} \int_{\frac{(i-1)T}{N}}^{\frac{iT}{N}} |\dot{p}(t)|\, dt \leqslant \alpha_1$. Therefore $s \in P_N^{1,\infty}$.

Let us consider the transportation map coupling the curve arcs between times $(i-1)\frac{T}{N}$ and $i\frac{T}{N}$ and the Diracs at $p\left(i\frac{T}{N}\right)$. Then

$$\begin{aligned} W_1(p_*\gamma, s_*\gamma) &\leqslant \sum_{i=1}^{N} \frac{1}{N} \sup_{(i-1)\frac{T}{N} \leqslant t \leqslant i\frac{T}{N}} \left\| s(t) - s\left((i-1)\frac{T}{N}\right) \right\| \\ &\leqslant \alpha_1 \frac{T}{N}. \end{aligned}$$

2. Let us fix $\mu \in \mathcal{M}\left(P_N^{1,\infty}\right)$ and let $s \in P_N^{1,\infty}$ such that $s_*\gamma = \mu$. We set $p(0) = s(1)$, and:

$$p(t) = \begin{cases} s(1) & \text{for } t \in \left]0, \frac{T}{N}\right], \\ s(i) + \left(\frac{t}{\Delta t} - i\right)(s(i+1) - s(i)) & \text{for } t \in \left]\frac{iT}{N}, \frac{(i+1)T}{N}\right] \\ & , i \in \{1, \dots N-1\}. \end{cases}$$

Since $s \in \Omega^N$ and $\Omega$ is convex, $p([0,T]) \subset \Omega$. Moreover, $p$ is continuous and piecewise differentiable. Finally, for $i \in \{1, \dots, N-1\}$ and $t \in \left]\frac{iT}{N}, \frac{(i+1)T}{N}\right]$, $\dot{p}(t) = \frac{1}{\Delta t}(s(i+1) - s(i)) = D_1(s)(i)$. Therefore, $\|\dot{p}\|_\infty \leqslant \alpha_1$, ensuring that $p \in \mathcal{P}_T^{1,\infty}$. With the same coupling as above, we have $W_1(p_*\gamma, s_*\gamma) \leqslant \alpha_1 \frac{T}{N}$, which ends the proof.

□

To end up, let us describe precisely a solver for the following variational problem:

$$\inf_{\mu \in \mathcal{M}\left(\mathcal{P}_T^{1,\infty}\right)} \mathcal{N}_h(\mu - \pi). \tag{5.30}$$

We let $\mathcal{M}^*$ denote the set of minimizers and $\mathcal{P}^*$ denote the associated set of parameterizations.

**Algorithm 3:** A projection algorithm on $\mathcal{M}\left(\mathcal{P}_T^{1,\infty}\right)$.

**Input**:

- $\pi$: target measure.
- $N$: a number of discretization points.
- $s^{(0)} \in P_N^{1,\infty}$: initial parameterized curve.
- $H$: a semi-algebraic function with Lipschitz continuous gradient.
- $nit$: number of iterations.

**Output**:

- $s^{(nit)}$: an approximation of a curve in $\mathcal{P}^*$.
- $\mu^{(nit)} = (s^{(nit)})_*\gamma_T$: an approximation of an element of $\mathcal{M}^*$.

**for** $0 \leq k \leq nit$ **do**

- Compute $\eta^{(k)} = \nabla J(s^{(k)})$
- Set $s^{(k+1)} = P_{P_N^{1,\infty}}\left(s^{(k)} - \tau\eta^{(k)}\right)$

*Remark* 5.27. The implementation of Algorithm 3 requires computing the gradients (5.21) and (5.22) and computing a projection on $P_N^{1,\infty}$. Both problems are actually non trivial.

The naive approach to compute the gradient of $F$ consists of using the explicit formula (5.21). This approach is feasible only for a small amount of points $N$ (less than 1000) since its complexity is $\mathcal{O}\left(N^2\right)$. In our numerical experiments, we therefore resort to fast summation algorithms (Potts and Steidl, 2003; Keiner et al., 2009) commonly used in particles simulation. This part of the numerical analysis is described in (Teuber et al., 2011) and we do not discuss it in this paper.

The set $P_N^{1,\infty}$ and more generally the sets $P_N^{m,q}$ are convex for $q \in [1, \infty]$. Projections can be computed using first-order iterative algorithms for convex functions. In our numerical experiments, we use accelerated proximal gradient descents on the dual problem (Beck and Teboulle, 2009a; Nesterov, 2013; Weiss et al., 2009). A precise description is given in (Chauffert et al., 2014b).

## 5.6 Results

To illustrate the results, we focus on the continuous line drawing problem discussed throughout the paper. It is performed using Algorithm 3. In the following experiments, we set $H$ as the opposite of a smoothed $L^2$-norm. This is similar to what was proposed in the original halftoning papers in (Schmaltz et al., 2010; Teuber et al., 2011).

### 5.6.1 Projection onto $P_N^{1,\infty}$

In this part, we limit ourselves to the projection onto $P_N^{1,\infty}$ as studied in the previous section. In Figure 5.3, we show the evolution of the curve $s^{(k)}$ across iterations, for different choices of $s^{(0)}$. After $30,000$ iterations, the evolution seems to be stabilized. The cost function during the 400 first iterations is depicted in Figure 5.4 for the three different initializations.

In Figure 5.5, we show the projection of the famous Girl with a Pearl Earring painting, after $10,000$ iterations. To really see the precision of the algorithm, we advise the reader to blink the eyes or to take a printed version of the paper away. From a close distance, the curves or points are visible. From a long distance, only the painting appears.

### 5.6.2 Projection onto $P_N^{m,q}$

We now consider projections onto more general measure spaces, such as $\mathcal{M}\left(\mathcal{P}_T^{m,q}\right)$, in order to show that different measures spaces can be considered. In Fig. 5.6, we show different behaviours for different $m \in \{1,2\}$ and $q \in \{1,2,\infty\}$. We also show a large scale example with a picture of Marylin Monroe in Figure 5.7.

## 5.7 Conclusion

We analyzed the basic properties of a variational problem to project a target Radon measure $\pi$ on arbitrary measures sets $\mathcal{M}_N$. We then proposed a numerical algorithm to find approximate solutions of this problem and gave several guarantees. An important application covered by this algorithm is the projection on the set of $N$-point measures, which is often called quantization and appears in many different areas such as finance, imaging, biology,... To the best of our knowledge, the extension to *arbitrary* measures set is new, and opens many interesting application perspectives. As examples in imaging, let us mention open topics such as the detection of singularities (Aubert et al., 2005) (e.g. curves in 3D images) and sparse spike deconvolution in dimension $d$ (Duval and Peyré, 2013).

To finish, let us mention an important open question. We provided necessary and sufficient conditions on the sequence $(\mathcal{M}_N)_{N\in\mathbb{N}}$ for the sequence of *global* minimizers $(\mu_N^*)_{N\in\mathbb{N}}$ to weakly converge to $\pi$. In practice, finding the global minimizer is impossible and we can only expect finding critical points. One may therefore wonder whether all sequences of critical points weakly converge to $\pi$. An interesting perspective to answer this question is the use of mean-filed limits (Fornasier et al., 2013).

## Acknowledgements

The authors wish to thank Gabriele Steidl for a nice presentation on halftoning which motivated the authors to work on this topic. The authors wish to thank Daniel Potts, Toni Volkmer and Gabriele Steidl for their support and help to run the excellent NFFT library (Keiner et al., 2009). They wish to thank Pierre Emmanuel Godet and Chan Hwee Chong for authorizing them to use the pictures in Figure 5.2.

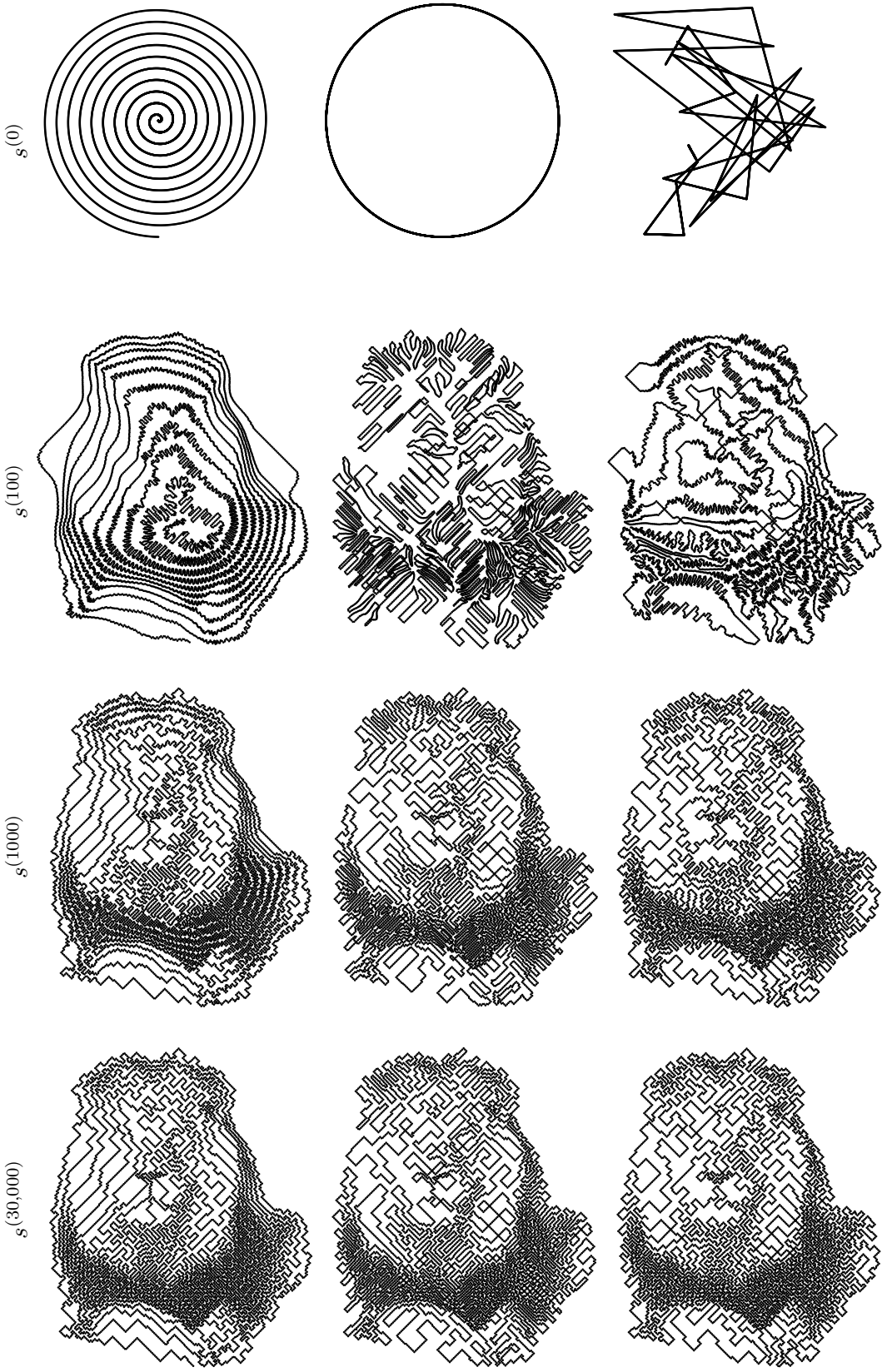


FIGURE 5.3: Projection of the lion image onto $P_N^{1,\infty}$ with $N = 8,000$. The figure depicts $s^{(k)}$ with several values of the iterate $k$ in Algorithm 3.

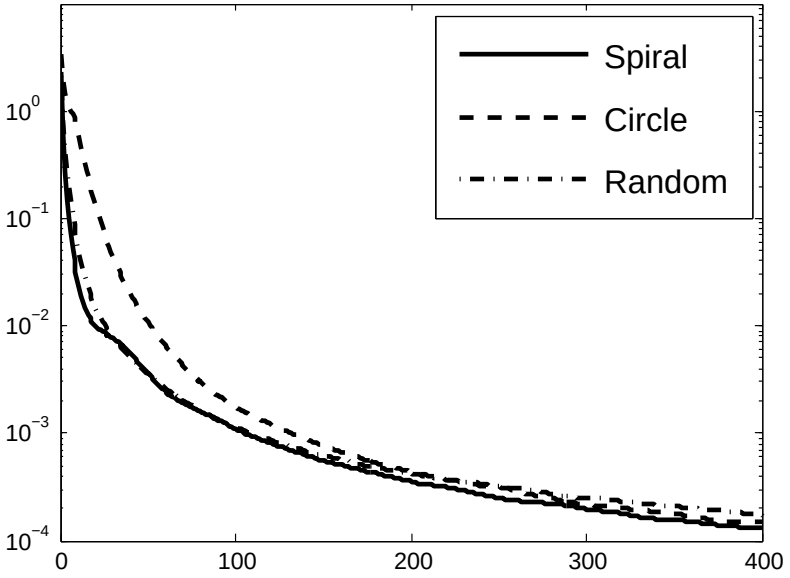


FIGURE 5.4: Decay of the cost function $J$ for the three experiments depicted in Fig. 5.3. We represent $\log_{10}(J(k) - m)$ for $k \leq 400$ where $m$ is the mimimal value of $J$ during the first $30,000$ iterations.

FIGURE 5.5: Projection of Meisje met de Parel, Vermeer 1665, onto $P_N^{1,\infty}$ with $N = 150,000$. The figure depicts $s^{(10,000)}$ obtained with Algorithm 3.

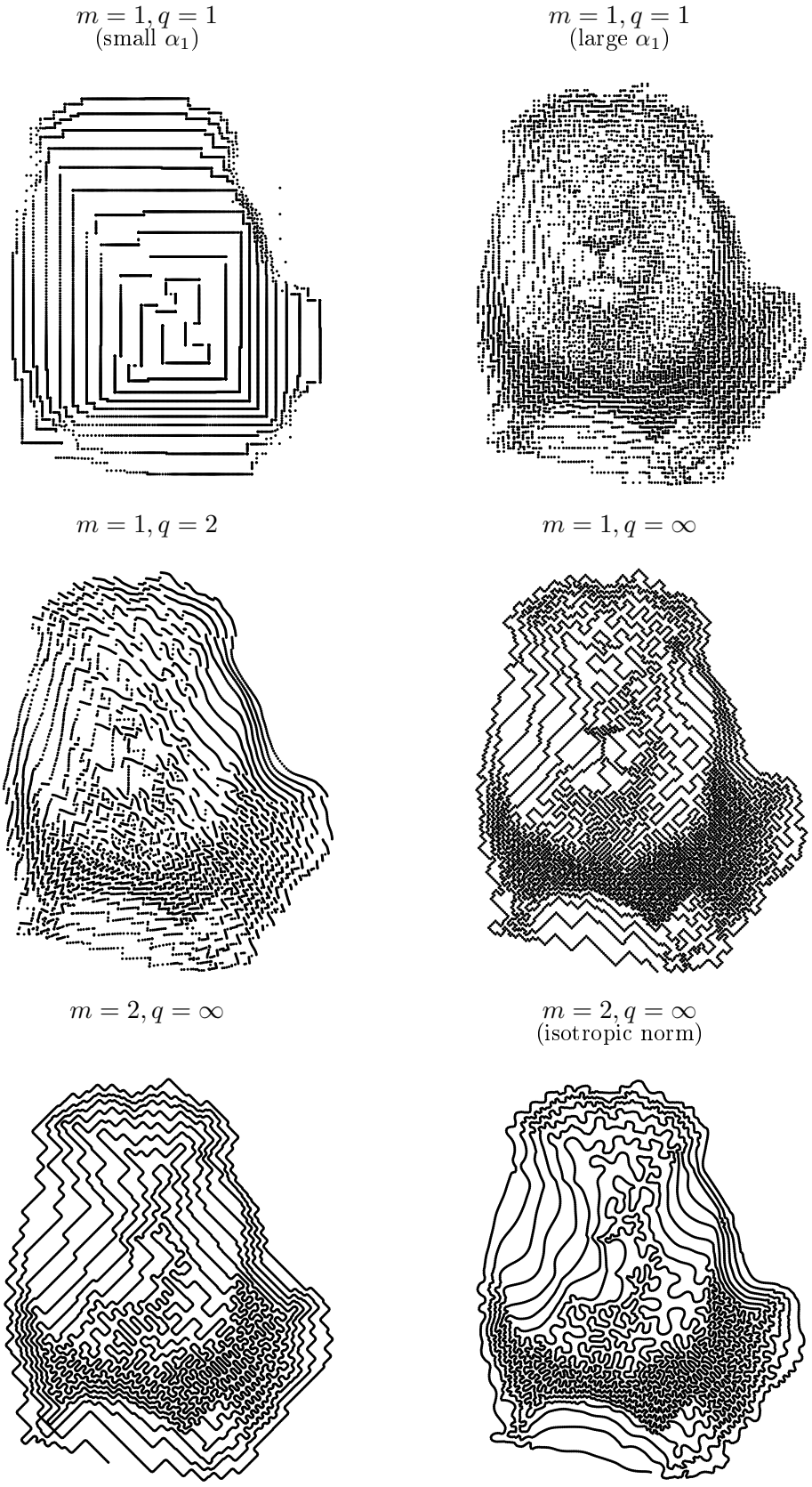


FIGURE 5.6: Projection of the lion image onto $P_N^{m,q}$ with $N = 8,000$, and $m \in \{1, 2\}$ and $q \in \{1, 2, \infty\}$.

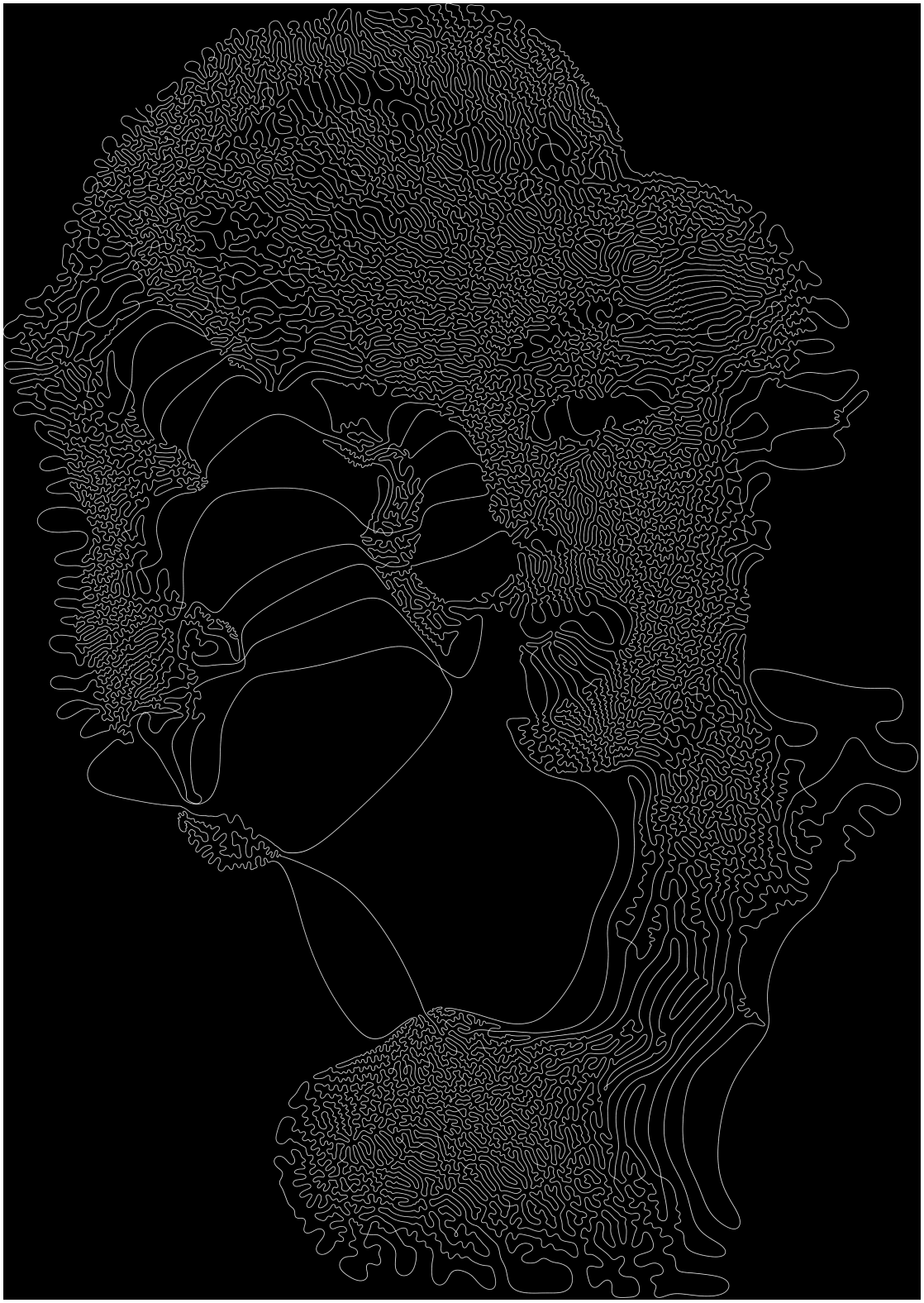

FIGURE 5.7: Projection of Marylin image, onto the set: $\mathcal{C} = \{p \in (W^{2,\infty}([0,T]))^2, \sup_{i\in[1,N]} (\|D_1 p(i)\|_2) \leq \alpha_1, \sup_{i\in[1,N]} (\|D_2 p(i)\|_2) \leq \alpha_2\}$, with $N = 100,000$. The figure depicts $s^{(10,000)}$ obtained with Algorithm 3.

# Chapter 6

# On the generation of sampling schemes for magnetic resonance imaging

Magnetic resonance imaging (MRI) is probably one of the most successful application fields of compressed sensing. Despite recent advances, there is still a large discrepancy between theories and actual applications. Overall, many important questions related to sampling theory remain open. In this paper, we attack one of them: given a set of sampling constraints (e.g. sampling Fourier coefficients along smooth curves), how to optimally design a sampling pattern? We first derive three key aspects that should be carefully designed by inspecting the literature, namely *admissibility*, *limit of the empirical measure* and *coverage speed*. To fulfill them jointly, we then propose an original approach which consists of projecting a probability distribution onto a set of admissible measures. The proposed algorithm allows to handle arbitrary constraints and then automatically generates efficient sampling patterns. In MRI, the images reconstructed with the proposed approach have a significantly higher SNR (2-3 dB) than those reconstructed using more standard sampling patterns (e.g. radial, spiral), both for mid and very high resolution imaging.

## 6.1 Introduction

Magnetic resonance imaging (MRI) is one of the flagship applications of compressed sensing (CS). The combination of CS and MRI initially appeared in (Lustig et al., 2007), very shortly after the seminal CS papers (Candès and Tao, 2006; Candès et al., 2006b; Donoho, 2006). However, the way CS was initially implemented on real scanners strongly

departed from theory. Despite having no solid theoretical foundations, it proved useful in practice and triggered a massive interest both in the MRI and mathematics communities.

Since then, many researchers have tried improving the way CS-MRI is implemented. These attempts can be divided into two distinct tracks:

- The first one consists of improving the incoherence of the sensing basis by using techniques termed phase scrambling or spread spectrum and originally proposed by (Haldar et al., 2011; Puy et al., 2012a). This can be implemented using specific radio-frequency pulses (Haldar et al., 2011) or shim-coils (Puy et al., 2012a). A few available theories support these techniques (Romberg, 2009; Puy et al., 2012b).
- The second one consists of keeping the sensing basis unchanged: images are acquired by collecting Fourier samples and assuming sparsity in a wavelet basis. The problem then reformulates as the design of new sampling patterns. Examples ranging in this second category include patterns made of parallel lines (Lustig et al., 2007), radial lines (Winkelmann et al., 2007), spirals (Nishimura et al., 1995), noisy spirals (Lustig et al., 2005), Poisson disc sampling (Vasanawala et al., 2011), ... Despite recent progresses, solid theoretical foundations for those approaches are still lacking.

To the best of our knowledge, and even though no report formally compared both approaches, the second is adopted more widely and provides a more efficient under-sampling and thus a faster acquisition in practice. A few numerical simulations to illustrate this fact were recently proposed in (Roman et al., 2014). In this paper, we will therefore concentrate on the second approach.

Sampling patterns proposed in the literature may seem somewhat arbitrary. For instance, even though existing theories recommand using completely random sampling patterns, it is not clear that adding random perturbations to a spiral will improve its practical efficiency.

**Contributions.** The first contribution of this paper is to provide a review of existing theoretical CS results in Section 6.3. This review permits to establish general principles for designing efficient sampling patterns.

The second and most significant contribution is to provide a constructive algorithm that generates feasible sampling patterns complying with the proposed principles in Sections 6.4–6.5. The main idea is to project a probability distribution onto a space of admissible measures. The reader can look at the result on Figure 6.1 to get an idea

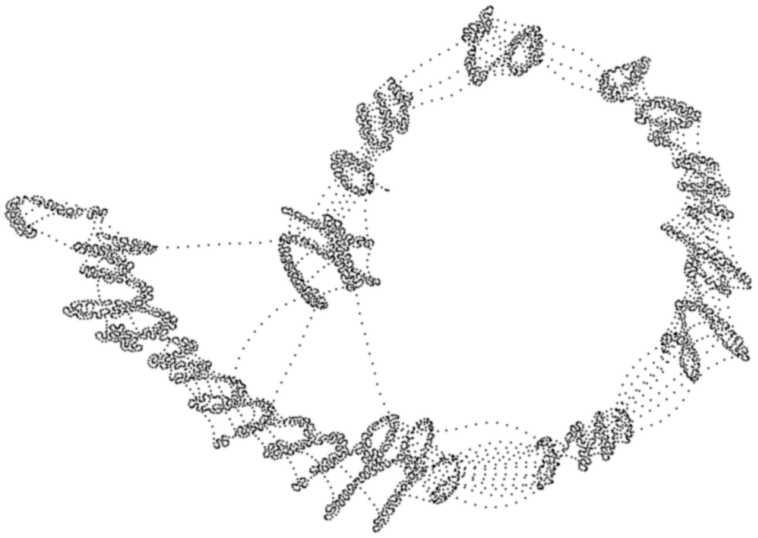

FIGURE 6.1: A glance at our contribution: our algorithm generates a sampling pattern complying with the MRI scanner constraints in which sampling locations consist of a piece of text, namely *How to sample me efficiently?*.

of what the algorithm does: given an initial distribution (here a piece of text), the algorithm finds a sampling pattern complying with physical constraints that best fits the distribution. A few experiments led on low and high resolution images suggest that the proposed sampling patterns significantly outperform more traditional approaches.

**Related works.** A few works in the literature address the problem of optimizing the acquisition space coverage using computational techniques.

The works (Mir et al., 2004; Spiniak et al., 2005) propose an algorithm to cover the whole $k$-space as fast as possible by using techniques used for missile guidance. This idea departs from the proposed one since the objective of these authors was to satisfy Shannon's sampling theorem, meaning that the samples should cover the space uniformly.

In (Kumar Anand et al., 2008; Curtis and Anand, 2008), the authors have proposed to synthetize random feasible trajectories using optimization techniques. Their idea was to generate random control points uniformly distributed over the surface of a sphere. They then searched for a feasible trajectory that passed close to them using second order cone programming. Multiple random trajectories were then generated this way and a

genetic algorithm was involved to select the most relevant ones so as to ensure a uniform $k$-space coverage. This idea does not stem from a clear sampling theory and is based on randomness, contrarily to the approach proposed in our paper.

Finally, a few authors (Seeger et al., 2010; Ravishankar and Bresler, 2011a; Liu et al., 2012) have borrowed ideas from statistical design for generating efficient sampling trajectories. In (Seeger et al., 2010), the key point is to fix a set of feasible trajectories (e.g. pieces of spirals) and to select them iteratively by picking the one that brings the largest amount of information at each step. Hence, finding the most meaningful trajectory becomes computationnaly intensive and hardly compatible with a real-time acquisition. The main contribution of (Ravishankar and Bresler, 2011a; Liu et al., 2012) is to propose alternative approaches to reduce the computational burden, by working on training images. These adaptive approaches suffer from a few drawbacks. First, the whole versatility of the MRI scanner is not exploited since fixed trajectories are imposed. Our formalism does not impose such a restriction. Second, even though adaptivity to the sampled image may seem appealing at the first glance, it still seems unclear whether this learning step is really helpful (Arias-Castro et al., 2013). Last but not least, these approaches strongly depart from existing sampling theories, while our contribution, though heuristic, is still motivated by solid and very recently established theories.

**Outline of the paper.** We first briefly describe the way images are acquired and reconstructed in MRI in Section 6.2. We then propose a short review of theoretical compressed sensing results in Section 6.3. Section 6.4 describes the main idea of the paper: we explain how the design of sampling patterns can be formulated as a measure projection problem. We then develop a numerical algorithm to solve this projection problem in Section 6.5. Finally, numerical experiments in a retrospective CS framework are conducted in Section 6.6 and conclusions are drawn in Section 6.7.

## 6.2 Acquisition and reconstruction in MRI

In this section, we first recall how images are sampled in MRI. We then describe standard reconstruction methods based on least squares or $\ell^1$-norm reconstructions using regridding techniques or non-uniform fast Fourier transforms.

### 6.2.1 Acquisition in MRI

In MRI, images are usually sampled in the so-called $k$-space domain, which corresponds to the 2D or 3D Fourier domain (Twieg, 1983). The acquisition domain can be slightly

different (i) in the parallel MRI context where spatial sensitivity encoding associated with the multiple channel coil introduces a convolution in $k$-space (Sodickson and Manning, 1997; Pruessmann et al., 1999) or (ii) when shim coils (e.g. phase scrambling/spread spectrum) are involved (Maudsley, 1988; Haldar et al., 2011; Puy et al., 2012a). In this paper, we focus on the Fourier domain, but the proposed ideas could be extended to these other settings.

The samples lie along parameterized curves $s : [0, T] \mapsto \mathbb{R}^d$, where $d \in \{2, 3\}$ denotes the image dimensions. The $i$-th coordinate of $s$ is denoted $s_i$. Let $u : \mathbb{R}^d \to \mathbb{C}$ denote a $d$ dimensional image and $\hat{u}$ be its Fourier transform. Given an image $u$, a curve $s : [0, T] \to \mathbb{R}^d$ and a sampling period $\Delta t$, the image $u$ shall be reconstructed from the following dataset:

$$\mathcal{E} = \left\{ \hat{u}(s(j\Delta t)), 0 \leqslant j \leqslant \left\lfloor \frac{T}{\Delta t} \right\rfloor \right\}. \tag{6.1}$$

In what follows, the scalar $m = \lfloor \frac{T}{\Delta t} \rfloor + 1$ denotes the total number of collected samples. Vector $\mathbf{y}$ with components $\mathbf{y}_j = \hat{u}(s(j\Delta t))$ denotes the vector of measurements. In this paper, we neglect typical distortions occurring in MRI such as noise, geometric distortions, signal loss at tissue/air interfaces or off-resonance effects which would strongly affect the dataset in Equation (6.1). We also neglect imprecisions in the trajectory due to Eddy currents that induce gradient errors (Brodsky et al., 2009). These are very important features that we plan to consider in forthcoming works.

The gradient waveform associated with a curve $s$ is defined by $g(t) = \gamma^{-1}\dot{s}(t)$, where $\gamma$ denotes the gyro-magnetic ratio (Hargreaves et al., 2004). The gradient waveform is obtained by supplying electric power to gradient coils. This electric current has a bounded amplitude and cannot vary too rapidly (slew rate). Mathematically, these constraints read:

$$\|g\| \leqslant G_{\max} \qquad \text{and} \qquad \|\dot{g}\| \leqslant S_{\max}$$

where $\|\cdot\|$ denotes either the $\ell^\infty$-norm defined by

$$\|f\|_\infty := \max_{1 \leq i \leq d} \sup_{t \in [0,T]} |f_i(t)|,$$

or the $\ell^{\infty,2}$-norm defined by

$$\|f\|_{\infty,2} := \sup_{t \in [0,T]} \left( \sum_{i=1}^{d} |f_i(t)|^2 \right)^{\frac{1}{2}}.$$

Additional affine constraints could be added depending on the targeted application (e.g.

structural or functional imaging) and the chronogram of the sequence (i.e. the interplay between the orthogonal gradients). For instance, $s$ usually starts from the $k$-space center, i.e. $s(0) = 0$. Multiple sampling trajectories (or interleaves) starting from the origin can be used to improve the signal-to-noise ratio: this typically leads to additional linear constraints of type $s(k \cdot TR) = 0$ for all $k \in \mathbb{N}$, where $TR$ is the time of repetition. Overall these additional constraints can be summarized under the compact form $A(s) = b$ where $A$ is a linear mapping and $b$ is a fixed vector. We refer to (Hargreaves et al., 2004; Chauffert et al., 2014b) for a more thorough discussion on these issues.

A sampling trajectory $s : [0, T] \to \mathbb{R}^d$ will be said *admissible* if it belongs to the set:

$$\mathcal{S}_T := \left\{ s \in \left(\mathcal{C}^2([0,T])\right)^d, \|\dot{s}\| \leqslant \alpha, \|\ddot{s}\| \leqslant \beta, A(s) = b \right\}. \tag{6.2}$$

In addition to the above mentioned kinematics constraints, important considerations regarding the MR signal acquisition have to be taken into account. The MR signal measures the amount of transversal relaxation, which exponentially decays as $\exp(-t/T_2)$, where $T_2$ is the transverse relaxation parameter. At the same time, the longitudinal relaxation begins to recover at a speed proportional to $(1 - \exp(-t/T_1))$ where $T_1$ is the longitudinal relaxation time. Both $T_1$ and $T_2$ are tissue specific. Two acquisition parameters $(TE, TR)$ permit to generate different weighted images (see (Brown et al., 2014) for details) depending on the choice for this pair. The echo-time $TE$ corresponds to the timing where the echo is generated, hence where the signal level is maximal whereas the $TR$ parameter reflects the duration between two consecutive RF pulses. Here, we will consider that the MR signal is available for about 200 ms. Therefore, the sampling time along one 2D trajectory should not exceed 200 ms. This requires choosing a long $TE$ of about 110 ms around which the readout of the signal will be performed by any trajectory (eg, from 10 ms to 210 ms). From a physical point of view, our numerical experiments will target $T_2$-weighted (long $TE$) imaging[1]. Hence, $TR$ should be chosen long too to unweight the contrast image from any longitudinal relaxation component. This constraint is not stringent in practice since 3D imaging is performed by iterating over slices. Hence, the $TR$ value for the whole volume corresponds to the number of slices multiplied by 2D acquisition time.

The last supplementary constraint is the maximal number of samples that can be stored in the buffer of the analogic-to-digital converter. This buffer length may depend on the imaging device but here we set this constraint to 8192 samples per readout.

[1] To be more accurate, $T_2^*$-weighted are obtained when involving gradient echo sequences.

### 6.2.2 Reconstruction in MRI

Reconstruction of MRI images from $k$-space measurements $\mathcal{E}$ is an involved problem that has been studied thoroughly. The main technical difficulties to solve it are (i) the fact that $k$-space locations $s(j\Delta t)$ do not lie on a Cartesian grid, (ii) the ill-posedness of the problem, (iii) the large image dimensions and (iv) an inaccurate knowledge of the acquisition operator owing to magnetic field inhomogeneities, subject movements,... The aim of this paragraph is to recall the main techniques developed so far to solve problems (i), (ii) and (iii). Altough of primary importance for implementing new sampling designs on scanners, we do not discuss problem (iv) since it is beyond the scope of this paper.

#### 6.2.2.1 Regridding or nonuniform fast Fourier transforms

In practice, the locations of Fourier samples $s(j \cdot \Delta t)$ seldom lie on a Cartesian grid. Standard discrete Fourier transforms can therefore not be used. To handle this situation, two strategies have been devised: regridding or non-uniform Fourier transforms.

Regridding techniques (see eg (Jackson et al., 1991; O'sullivan, 1985)) are probably the most widespread techniques. They consist of interpolating the information lying on available arbitrary locations to positions lying on a Cartesian grid. The basic idea is to convolve the non-uniform Fourier samples with a regularizing kernel (e.g. Kaiser-Bessel) and to resample the result.

In this paper we will use a less common approach based on Non-Uniform Fast Fourier transforms (NUFFT) (Knopp et al., 2007). This approach was actually shown to be equivalent to regridding techniques with a Gaussian kernel (Sarty et al., 2001). It presents the advantage of coming with good parallel implementations on multi-core or GPU architectures (Keiner et al., 2009; Freiberger et al., 2013).

#### 6.2.2.2 Regularization

In cases where the whole Fourier domain is sampled on a sufficiently fine Cartesian grid, the reconstruction problem is well posed, in the sense that there exists a unique image that explains the measurements. Moreover, this image can be reconstructed in a stable manner by simply inverting the sensing matrix. This can be done using the fast Fourier transform.

In contrast, when the samples are not located on a Cartesian grid and/or if each slice of the image to be reconstructed contains more pixels than the number of collected samples, the problem becomes ill-posed and a direct inversion of the sensing matrix is impossible.

To solve it, various strategies have been developed. Some of them are briefly discussed hereafter.

**Regridding and inversion.** One of the most standard techniques to reconstruct MRI images consists of first regridding the non-uniform samples on a Cartesian grid and performing an inverse Fast Fourier transform to recover an image. This technique provides good results when the samples are sufficiently dense. It performs very poorly for strongly undersampled data.

**Reconstruction using least squares.** To regularize the problem, another standard technique consists of using least squares or Tikhonov regularization. Let $\mathbf{S} : \mathbb{C}^n \to \mathbb{C}^m$ denote the linear operator that maps the discrete image to its Fourier transform values at locations $s(j \cdot \Delta t)$. Matrix-vector products with this operator can be computed using regridding techniques or NUFFT. The least squares formulation consists of finding the minimizer of $\|\mathbf{Su} - \mathbf{y}\|_2$, where $\mathbf{u} \in \mathbb{C}^n$ denotes the discrete image to reconstruct. Tikhonov regularization is then incorporated for minimizing the following penalized criterion:

$$\min_{\mathbf{u}\in\mathbb{R}^n} \frac{1}{2}\|\mathbf{Su} - \mathbf{y}\|_2^2 + \frac{\lambda}{2}\|\mathbf{Du}\|_2^2$$

where $\mathbf{D}$ is a matrix that defines the regularizer (e.g. identity or finite differences) and $\lambda \in \mathbb{R}_+$ is the regularization parameter.

Both the least squares and Tikhonov regularization can be solved very efficiently using simple (preconditioned) conjugate gradient descents. This feature explains their success. It is however now well known that better results can be obtained using non-linear programming.

**Reconstruction using $\ell^1$-norms.** The theory of compressed sensing triggered a massive interest in the use of the sparsity promoting $\ell^1$-norm regularization. We will review some of its theoretical guarantees in Section 6.3. The idea is to decompose the image $\mathbf{u}$ on a basis or a frame $\mathbf{\Psi} \in \mathbb{R}^{p\times n}$, where $p \geq n$ denotes the number of atoms in the frame. In matrix notation, the decomposition reads $\mathbf{u} = \mathbf{\Psi x}$ where $\mathbf{x} \in \mathbb{R}^p$ denotes the coefficients of $\mathbf{u}$ in the frame. For a basis, $p = n$ and the decomposition $\mathbf{x}$ is unique. For a redundant frame ($p > n$) there is an infinity of decompositions in a frame. It is well known that many bases or frames such as wavelets, curvelets or shearlets allow compressing the information present in $\mathbf{u}$, meaning that among all decompositions, there exists at least one of kind $\mathbf{u} = \mathbf{\Psi x}$ such that most of the energy of $\mathbf{x}$ is concentrated in a small number of nonzero coefficients.

This observation motivated the introduction of the basis pursuit algorithm that consists of solving:

$$\min_{\mathbf{x}\in\mathbb{R}^p, \mathbf{S\Psi x}=\mathbf{y}} \|\mathbf{x}\|_1. \tag{6.3}$$

The use of the $\ell^1$-norm is often justified as a convex relaxation of the $\ell^0$-counting function, that counts the number of nonzero components in $\mathbf{x}$. When the data $\mathbf{y}$ is degraded by noise, the exact constraint $\mathbf{S\Psi x} = \mathbf{y}$ is relaxed and transformed into a penalized data consistency term. Then, the following quadratic programming problem has to be solved instead:

$$\min_{\mathbf{x}\in\mathbb{R}^p} \|\mathbf{x}\|_1 + \frac{\lambda}{2}\|\mathbf{S\Psi x} - \mathbf{y}\|_2^2. \tag{6.4}$$

Scalar $\lambda > 0$ is a parameter that balances the quadratic data consistency term and the regularization term. In this paper, we will mainly use this last formulation, since to the best of our knowledge, it is the one associated with the strongest theoretical reconstruction guarantees. In all the paper, $\mathbf{\Psi}$ is defined as an orthogonal wavelet transform with Daubechies wavelets and 4 vanishing moments. Therefore $p = n$.

Problem (6.4) can be solved by using various well documented techniques. In this paper we will use an accelerated proximal gradient descent algorithm (aka FISTA) (Nesterov, 1983; Beck and Teboulle, 2009b).

**More advanced reconstruction techniques.** Finally, let us acknowledge that the most efficient reconstruction strategies do not rely on a simple $\ell^1$-minimization as described in the last paragraph. More advanced regularizers are usually more effective. One possibility is to use analysis prior as regularizers such as total variation (Block et al., 2007). One of the most popular approaches currently consists of combining analysis and synthesis priors. The idea is to use an objective function of type:

$$\min_{\mathbf{x}\in\mathbb{R}^p} \gamma\|\mathbf{D\Psi x}\|_1 + \|\mathbf{x}\|_1 + \frac{\lambda}{2}\|\mathbf{S\Psi x} - \mathbf{y}\|_2^2, \tag{6.5}$$

where $\mathbf{D}$ is a matrix that may represent differential or time-frequency operators (Ma et al., 2008; Boyer et al., 2012). Another recent trend consists of learning the representation dictionary (i.e. $\mathbf{\Psi}$) rather than fixing it in advance (Ravishankar and Bresler, 2011b).

Overall, these methods provide more competitive alternatives to the simple $\ell^1$-reconstruction method (6.4). We do not use them in this paper for three reasons. First, solving the optimization problems arising with such approaches is usually more computationally demanding. Second, it is often hard to set the additional regularization parameters properly, making numerical tests much more complicated. Finally, the most efficient image reconstructors have so far few theoretical reconstruction guarantees.

#### 6.2.2.3 Wavelet crime

In this paper, we consider a discrete problem, meaning that we evaluate Fourier transform values of a discrete signal.

On a real MRI scanner, the acquired data come from a continuous signal which is discretized by the analogous-to-digital converter. Many authors commit what is commonly referred to as the wavelet crime: one implicitely assumes that the signal is discrete whereas it is actually continuous (see (Strang and Nguyen, 1996)). This usually leads to severe ringing artifacts in the reconstruction. Solutions to this problem have been proposed in many works. We refer to eg (Guerquin-Kern et al., 2011; Adcock and Hansen, 2011) for a more thorough description of these methods.

#### 6.2.2.4 Parallelization

One of the main difficulties in MRI image reconstruction lies in the high dimensionality of images. This usually leads to very long computing times that may be incompatible with clinical routine. Many authors recently made use of the progresses in computers and multicore programming to accelerate their reconstructions.

Some are based on standard multi-core architectures using e.g. OpenMP (Murphy et al., 2012; Chang and Ji, 2010). Others are based on the more recent GPU parallelization (Smith et al., 2012).

In this paper, all the numerical experiments are based on the NUFFT3 package delivered by Chemnitz university (Keiner et al., 2009). This library is natively parallel.

## 6.3 Theoretical foundations of variable density sampling

In this section, we briefly review the existing theoretical CS results. The conclusions of this section motivate the main contribution of this work: the design of undersampling patterns by measure projection.

### 6.3.1 The first compressed sensing results

Let us first describe the compressed sensing theory as it appeared in the seminal paper (Candès et al., 2006b) and more recently in (Candès and Plan, 2011). The authors

consider an orthogonal matrix

$$\mathbf{A}_0 = \begin{pmatrix} \mathbf{a}_1^* \\ \vdots \\ \mathbf{a}_n^* \end{pmatrix}.$$

They propose to construct a random sensing matrix as:

$$\mathbf{A} = \begin{pmatrix} \mathbf{a}_{J_1}^* \\ \vdots \\ \mathbf{a}_{J_m}^* \end{pmatrix},$$

where the integers $J_k \in \{1, \dots, n\}$ are i.i.d. uniform random variables.

Knowing that $\mathbf{y} = \mathbf{A}\mathbf{x}$ the authors propose to recover $\mathbf{x}$ by solving Problem (6.3). Let

$$\bar{\mathbf{x}} = \underset{\mathbf{x} \in \mathbb{R}^p, \mathbf{A}\mathbf{x} = \mathbf{y}}{\arg\min} \|\mathbf{x}\|_1.$$

Their main result in the noiseless case reads as follows:

**Theorem 6.1.** *Assume that* $\mathbf{x}$ *is* $s$*-sparse, i.e. that it contains at most* $s$ *nonzero components. If the number of measurements* $m$ *satisfies:*

$$m \geq Cs \left( n \max_{1 \leq k \leq n} \|\mathbf{a}_k\|_\infty^2 \right) \log\left(\frac{n}{\epsilon}\right),$$

*where* $C$ *is a universal constant, then* $\bar{\mathbf{x}} = \mathbf{x}$ *with probability* $1 - \epsilon$*.*

Moreover, the authors show that if the measurements are noisy, i.e. $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{b}$, where $\mathbf{b}$ is a random perturbation, then the solution to the relaxed Problem (6.4) also provides stable reconstruction results.

The *coherence* $\kappa(\mathbf{A}_0) = n \max_{1 \leq k \leq n} \|\mathbf{a}_k\|_\infty^2$ belongs to the interval $[1, n]$. In particular, $\kappa(\mathbf{F}) = 1$ and $\kappa(I_d) = n$. In the favorable case of a Fourier transform, this theorem indicates that only $s \log\left(\frac{n}{\epsilon}\right)$ measurements are enough to perfectly recover an arbitrary $s$-sparse signal.

Even though this type of theorem got a huge impact in the literature, it is not applicable to MRI. The natural transform $\mathbf{A}_0$ in MRI reads $\mathbf{A}_0 = \mathbf{F}^* \mathbf{\Psi}$, i.e. the product of Fourier and wavelet transforms. In that case, one can show that $\kappa(\mathbf{A}_0) = O(n)$. Theorem 6.1 is thus irrelevant in such a setting.

### 6.3.2 The emergence of variable density sampling

In most practical applications, the transforms $\mathbf{A}_0$ are coherent. This is the case in MRI and more generally in Fourier or space imaging. A simple technique to break the so-called "coherence barrier" consists of drawing the coherent samples more often than the ones with low coherence (Puy et al., 2011; Krahmer and Ward, 2014; Chauffert et al., 2014a). Let us clarify this idea. Let $\boldsymbol{\pi} \in \Delta_n$ denote the distribution of the i.i.d. random variables $J_k$, i.e. $P(J_k = i) = \boldsymbol{\pi}_i$. The following theorem (Chauffert et al., 2014a) justifies the use of variable density sampling.

**Theorem 6.2.** *Assume that* $\mathbf{x}$ *is* $s$*-sparse, i.e. that it contains at most* $s$ *nonzero components. Set*

$$\boldsymbol{\pi}_k = \frac{\|\mathbf{a}_k\|_\infty^2}{\sum_{j=1}^n \|\mathbf{a}_j\|_\infty^2}.$$

*If the number of measurements satisfies*

$$m \geq Cs \left( \sum_{j=1}^n \|\mathbf{a}_j\|_\infty^2 \right) \log\left(\frac{n}{\epsilon}\right),$$

*where* $C$ *is a universal constant, then* $\bar{\mathbf{x}} = \mathbf{x}$ *with probability* $1 - \epsilon$.

One can show that in the case of MRI, $\sum_{j=1}^n \|\mathbf{a}_j\|_\infty^2 = O(\log(n))$. It is therefore possible to reconstruct exactly an $s$-sparse image with $O(s \log(n)^2)$ samples. Let us mention that variable density sampling was the basis for the seminal paper on compressed sensing MRI (Lustig et al., 2007). Theorem 6.2 is a first argument that supports that type of technique.

### 6.3.3 Variable density sampling with structured sparsity

Theorem 6.2 is quite attractive from a theoretical point of view. A simple analysis however suggests that it is still insufficient to justify the use of compressed sensing in MRI. First, the constant appearing in the $O$ is large. This may only be an artifact of the proofs, but it is currently unknown how much it can be lowered. More importantly, the term $\log(n)^2$ that appears when using the Fourier-Wavelet pair cannot be improved by using only variable density sampling arguments. Most often, the logarithmic terms are disregarded and considered as negligible. It is however important to look at them carefully, for instance $\log(1024 \times 1024)^2 = 192$. A method needing $192s$ samples to reconstruct a $1024 \times 1024$ image is of little practical interest.

A recent advance that seems very promising is proposed in (Adcock et al., 2013). The authors show that it is possible to exploit a *structured sparsity* to obtain better reconstruction guarantees. In the case of imaging, structured sparsity may mean that the wavelet subbands become sparser as the scale increases. Let us provide a typical result from this active field of research. This result is quite similar to (Adcock et al., 2013) and comes from a recent preprint (Boyer et al., 2015b).

Let $(\Omega_j)_{0\leq j\leq J}$ denote the wavelet subbands with $J$ the number of decomposition levels. Assume that $\mathbf{x}$ is supported on $S \subset \{1,\dots,n\}$ with $|S \cap \Omega_j| = s_j$. This means that $\mathbf{x}$ restricted to the subband $\Omega_j$ is $s_j$-sparse. This model is called sparsity by levels in (Adcock et al., 2013). In such a setting, we can prove the following theorem.

**Theorem 6.3.** *Assume that matrix $\mathbf{A}_0$ is the product of the Fourier and Haar Wavelet matrices. Let $j(k)$ denote the scale of index $k$, i.e. $j(k) = j$ if $k \in \Omega_j$. Set*

$$\boldsymbol{\pi}_k = \frac{2^{-j(k)} \sum_{p=0}^{J} 2^{-|j(k)-p|/2} s_p}{\gamma} \quad \textit{with} \quad \gamma = \sum_{j=0}^{J}\sum_{p=0}^{J} 2^{-|j-p|/2} s_p.$$

*Set*

$$m \geq C\gamma \log(s) \log\left(\frac{n}{\epsilon}\right) \tag{6.6}$$

*where $C$ is a universal constant.*

*Under the previous sparsity-by-level hypothesis $\bar{\mathbf{x}} = \mathbf{x}$ with probability $1-\epsilon$.*

Note that contrarily to previous results, the drawing probability $\boldsymbol{\pi}$ in Theorem 6.3 explicitly depends on the sparsity structure. The number of measurements in Theorem 6.3 is always lower than that of Theorem 6.2, but the gain once again depends on the signal support. At the price of extra technicalities (the weak-balancing property in (Adcock et al., 2013)), the term $\log(s)$ in Equation (6.6) can also be discarded.

### 6.3.4 Variable density sampling with structured acquisition

Another element that was not considered in the seminal works on compressed sensing is *structured acquisition*. In practice, sampling isolated measurements is not practical or even feasible. In MRI, radio-interferometry, X-ray tomography and many other systems, the samples have to lie on particular shapes or curves imposed by the physics of acquisition. The vast majority of compressed sampling schemes are based on heuristic sampling patterns such as radial lines (Lauterbur et al., 1973; Winkelmann et al., 2007), spirals (Spielman et al., 1995), noisy spirals (Vasanawala et al., 2011) or other exotic shapes.

Even though they often perform well, until very recently there were missing theoretical results that allow to justify their use in practice.

In the spirit of traditional Shannon's sampling theorem, the papers (Unnikrishnan and Vetterli, 2013; Gröchenig et al., 2014) propose theoretical guarantees for the reconstruction of bandlimited functions from sets of measurements along lines or curves. These results usually lead to sampling patterns that span the acquisition space uniformly.

Concomitantly to these developments, we have proposed a few results in (Bigot et al., 2013; Chauffert et al., 2014a; Boyer et al., 2015b) to explain the success of structured acquisitions by using sparsity assumptions on the signal to be reconstructed. These results promoted variable density sampling strategies. In (Bigot et al., 2013; Boyer et al., 2015b), theoretical guarantees were derived for block sampling strategies: instead of probing isolated measurements, fixed groups of measurements are acquired, irrespective of the structured sparsity assumptions. Still in these references, it is shown that only specific sparsity patterns that depend on the acquisition constraints can be recovered.

In (Chauffert et al., 2014a), we proposed to sample signals using generic stochastic processes. The conclusions of this work actually define the starting point of the present paper. We first gave a mathematical definition of variable density samplers as sequences of stochastic processes with a prescribed limit empirical measure, termed density. We also showed through mathematical arguments and experimental validation that the key features characterizing the efficiency of a variable density sampler are:

i) The *density*: the stochastic processes should cover the space non-uniformly according to a certain density.

ii) The *coverage speed*: a sampler will be efficient only if it covers the space fast enough. More precisely, we showed that the mixing time should be as low as possible. The mixing time characterizes the speed at which the empirical measure converges to its limit.

Since most readers may not be familiar with these concepts, we illustrate them in Figure 6.2. In this Figure, we constructed three different variable density samplers with a density $\boldsymbol{\pi}$ illustrated on Figure 6.2 (a). This density was defined as suggested by Theorem 6.2 by setting $\boldsymbol{\pi}_k \propto \|\mathbf{a}_k\|_\infty$, where $\mathbf{a}_k$ is the $k$-th row of the Fourier-Wavelet matrix $\mathbf{A} = \mathbf{F}^*\boldsymbol{\Psi}$. The wavelet transforms was defined using Daubechies 4 filters. The sampling schemes in Figure 6.2 (b,c,d) all cover the $256 \times 256$ grid non uniformly with 20% measurements. For the sampling patterns (b) and (d), the samples density in a given region of space looks like $\boldsymbol{\pi}$. It is also the same for (c) with a little bit of imagination. This property of

non uniform coverage is captured by the sampler's density (more precisely, the limit of the empirical measure), i.e. feature i).

It is pretty intuitive when looking at (b,c,d) that they are likely to have different efficiencies. The samples in Figure 6.2(b) cover the space quite uniformly locally, while the samplers in Figure 6.2(c)-(d) leave large portions of the space unexplored. Clearly, this lack of information might result in unsatisfactory reconstruction results. This feature is captured by the notion of coverage speed, i.e. feature ii). Let us mention that the so-called poisson disc sampling (Bridson, 2007; Murphy et al., 2012), which is quite popular in prospective compressed sensing MRI, is also based on the idea of covering the $k$-space as fast as possible.

## 6.4 Generation of sampling schemes by projection

In this section, we describe the main idea of this paper. We propose a general principle to construct samplers that comply with the three following rules:

- *Admissibility*: the sampler should be feasible, and therefore belong to a given set. For instance in the case of MRI, it can be a set of straight lines or the set of curves defined in Equation (6.2).
- *Density*: as mentioned earlier, a sampler should approximate a given density $\pi$.
- *Coverage speed*: the sampler should cover the space as fast as possible.

This problem is probably more complex than it looks at first sight. In this section we will first recall the notion of pushforward measure that is crucial to establish our algorithm. We then present its overall principle.

Let us mention that this idea, the associated algorithm and some of its theoretical guarantees were presented in more detail in our recent preprint (Chauffert et al., 2015a) for a completely different purpose, namely image stippling or continuous line drawing.

### 6.4.1 Pushforward measures

As can be seen in Figure 6.2, the density (a) is somehow similar to the sampling schemes (b, c and d). To make this statement more accurate, we resort to measure theory. Let us introduce a few definitions. Here, we work on the space $\Omega = [0,1]^d$ where $d = 2$ denotes the space dimension. Extensions to other dimensions are straightforward. We

equip $\Omega$ with the Borel algebra $\mathcal{B}$. Let $(X, \Sigma)$ be a measurable space, $f : X \to \Omega$ denote a measurable mapping and $\mu : X \to [0, +\infty]$ denote a measure. The *pushforward* measure of $\mu$ is denoted $\nu$ defined by

$$\nu(B) = f_*\mu(B) = \mu\left(f^{-1}(B)\right), \quad \forall B \in \mathcal{B}.$$

The function $f$ is called *parameterization* of $\nu$. Note that if $\mu$ is a probability measure, then $\nu$ is also a probability measure. Let us now illustrate this concept with two concrete examples.

**Atomic measures.** The set of $m$-points in Figure 6.2 (b) can be ordered and parameterized as a function $f : \{1, \dots, m\} \to \Omega$, where $f(i) = p_i$ denotes the $i$-th point. Set $\mu$ as the normalized counting measure defined for any set $I \subseteq \{1, \dots, m\}$ by $\mu(I) = \frac{|I|}{m}$. Let $B \in \mathcal{B}$, then $f^{-1}(B)$ is the set of indices of points in $B$. The pushforward of $\mu$ is therefore an atomic measure defined by

$$\nu = \mu_* f = \frac{1}{m} \sum_{i=1}^{m} \delta_{p_i}.$$

**Measures supported on curves.** Let $s : [0, T] \to \Omega$ denote a parameterized curve. Set $\mu$ as the normalized Lebesgue measure on $[0, T]$ defined for any interval $I \subseteq [0, T]$ by $\mu(I) = \frac{|I|}{T}$. Then $\nu(B) = s_*\mu(B)$ measures the relative time spent by the curve $s$ in the set $B$.

Now, let $\mathcal{P}$ denote a set of admissible parameterizations. We let $\mathcal{M}(\mathcal{P})$ the set of pushforward measures associated with elements of $\mathcal{P}$:

$$\mathcal{M}(\mathcal{P}) = \{\nu = f_*\mu, \ f \in \mathcal{P}\}.$$

Depending on the context, $\mu$ will be either the normalized counting measure or the normalized Lebesgue measure. Hereafter, we will be particularly interested in exploring 3 different sets $\mathcal{P}$ which are particularly relevant in MRI.

**Isolated points.** The set of sums of $m$ Dirac delta functions is:

$$\mathcal{M}(\Omega^m) = \left\{ \nu = \frac{1}{m} \sum_{i=1}^{m} \delta_{p_i}, \ p_i \in \Omega \right\}. \tag{6.7}$$

This case is not feasible or useful in MRI, but it is commonly used in simulations. It will therefore serve as a reference.

**Admissible curves for MRI.** It corresponds to $\mathcal{M}(\mathcal{S}_T)$, where $\mathcal{S}_T$ is defined by Equation (6.2). This case is the one that permits to exploit the full sampling potential in MRI.

**Lines of variable length.** Finally, we propose to use $N$ lines with variable lengths (or crossed with variable speed at constant time). To this end, we define:

$$\mathcal{L} = \{\lambda : [0,1] \to \Omega,\ \exists (x_1, x_2) \in \Omega^2,\ \lambda(t) = (1-t)x_1 + tx_2,\ \forall t \in [0,1]\}.$$

The associated set of measures is defined by:

$$\mathcal{M}(\mathcal{L}^N) = \left\{ \nu = \frac{1}{N} \sum_{i=1}^{N} (\lambda_i)_* \mu,\ \lambda_i \in \mathcal{L} \right\}, \tag{6.8}$$

where $\mu$ is the Lebesgue measure on $[0,1]$. This set of measures offers more versatility than the previous one and is quite simple to implement too. In this description, we implicitly assume that lines of different lengths are traversed at different speeds since the traversal time is fixed to 1.

### 6.4.2 Measuring distances between measures

Pushforward measures allow us to map a sampling pattern to the space of probability measures $\mathcal{M}_\Delta$. The target distribution $\pi$ also belongs to $\mathcal{M}_\Delta$. This mapping therefore enables to perform quantitative comparisons by defining distances on $\mathcal{M}_\Delta$. Various distances exist to compare probability measures (e.g. total variation, Wasserstein distance, ...). In this work, motivated by our previous results in (Chauffert et al., 2015a), we propose to construct a distance as follows. Let $h : \Omega \to \mathbb{R}$ denote a continuous function with a Fourier series that does not vanish. The following mapping:

$$\mathrm{dist}(\pi, \nu) = \|h \star (\pi - \nu)\|_2^2 \tag{6.9}$$

defines a distance (or metric) on $\mathcal{M}_\Delta$. Moreover, we showed in (Chauffert et al., 2015a) that it metrizes the weak convergence. Therefore, if $\pi$ and $\nu$ are sufficiently weakly close, their distance will be small.

This measure is interesting numerically for at least two reasons. First, it has a simple direct expression compared to more standard tools such as the Wasserstein distance. Second, it is quadratic and this property will be exploited intensively in the numerical algorithms.

### 6.4.3 Design of sampling scheme as a projection problem

The distance on $\mathcal{M}_\Delta$ being defined, we can construct a sampler by solving the following variational problem:

$$\min_{\nu \in \mathcal{M}(\mathcal{P})} \operatorname{dist}(\pi, \nu) \tag{6.10}$$

where $\mathcal{P}$ is the set of admissible parameterizations. In other words, we are looking for the admissible measure $\nu^*$ that is the closest to the target measure $\pi$. This is therefore a projection problem.

Let us mention that the mapping $\nu \mapsto \operatorname{dist}(\pi, \nu)$ is a nice convex and smooth function. However, for most parameterization sets $\mathcal{P}$, the associated measure set $\mathcal{M}(\mathcal{P})$ is highly nonconvex. This makes the resolution of Problem (6.10) very involved. In fact, in the "simple case" $\mathcal{P} = \Omega^m$, Problem (6.10) corresponds to Smale's 7th problem to solve for the XXIst century (Smale, 1998).

## 6.5 Numerical implementation

In this section, we describe a numerical algorithm to solve Problem (6.10). Starting from an initial user provided sampling pattern, the algorithm iteratively minimizes (6.9).

### 6.5.1 The attraction-repulsion formulation

In order to numerically solve Problem (6.9), we need to discretize it. It was shown in (Chauffert et al., 2015a) that any measure set $\mathcal{M}(\mathcal{P})$ can be approximated by a subset of $p$-point measures $\mathcal{N}_p \subseteq \mathcal{M}(\Omega^p)$ with an arbitrary precision. Precisely, it is possible to control their Hausdorff distance defined by:

$$\mathcal{H}_{\operatorname{dist}}(\mathcal{N}_p, \mathcal{M}(\mathcal{P})) = \max\left( \sup_{\pi \in \mathcal{N}_p} \inf_{\mu \in \mathcal{M}(\mathcal{P})} \operatorname{dist}(\mu, \pi), \sup_{\mu \in \mathcal{M}(\mathcal{P})} \inf_{\pi \in \mathcal{N}_p} \operatorname{dist}(\mu, \pi) \right).$$

Moreover, the set $\mathcal{N}_p$ can always be written as

$$\mathcal{N}_p = \left\{ \mu = \frac{1}{p} \sum_{i=1}^{p} \delta_{q_i}, \quad \text{for } q = (q_i)_{1 \leq i \leq p} \in \mathcal{Q}_p \right\},$$

where the parameterization set $\mathcal{Q}_p$ depends on $\mathcal{P}$. The abstract definition of $\mathcal{Q}_p$ proposed in (Chauffert et al., 2015a) is not constructive. Explicit constructions for the parameterizations considered hereafter are provided in the next section. Once an approximate

space of parameterizations $\mathcal{Q}_p$ is constructed, Problem (6.9) can be replaced by

$$\min_{\nu \in \mathcal{N}_p} \frac{1}{2} \|h \star (\nu - \pi)\|_2^2, \tag{6.11}$$

where $\mathcal{N}_p = \mathcal{M}(\mathcal{Q}_p)$ is a suitable approximation of $\mathcal{M}(\mathcal{P})$. Then, by developing the $L^2$-norm, we may rewrite Problem (6.11) as follows:

$$\min_{q \in \mathcal{Q}_p} \underbrace{\frac{1}{2} \sum_{i=1}^{p} \sum_{j=1}^{p} H(q_i - q_j)}_{J_1(q)} - \underbrace{\sum_{i=1}^{p} \int_{\Omega} H(x - q_i) d\pi(x)}_{J_2(q)}, \tag{6.12}$$

where $H$ is defined in the Fourier domain by $\widehat{H}(\xi) = |\widehat{h}|^2(\xi)$ for all $\xi \in \mathbb{Z}^d$. In this paper, we consider a kernel $H$ defined by $H(x) = -\|x\|_2$. It is rather easy to check that this kernel has a non-negative Fourier series and that it ensures rotation, translation and scale invariance of the global minimizers of (6.12).

The functional (6.12) can be decomposed in two terms:

- The first one $J_1$ is a *repulsion potential*: it tends to maximize the distance between all point pairs. It will ensure that no cluster of points emerges and therefore ensures a good space coverage.
- The second one $J_2$ is an *attraction potential*: it attracts the particles $q_i$ in the regions of high density of $\pi$. This term ensures that the solution of Problem (6.12) will have an appropriate density.

Let us note that the attraction-repulsion functional (6.12) was initially proposed in (Schmaltz et al., 2010; Teuber et al., 2011) as an alternative to Poisson disk sampling (Bridson, 2007; Vasanawala et al., 2011). The idea we propose can therefore be considered as a generalization of Poisson disk sampling, allowing to handle arbitrary additional constraints.

### 6.5.2 Projected gradient descents

The attraction-repulsion formulation (6.12) of the projection problem (6.10) is easily amenable to a numerical resolution. Similarly to (Chauffert et al., 2015a), we propose to use a projected gradient descent. We only describe it briefly hereafter and refer to (Chauffert et al., 2015a) for its theoretical guarantees and more details. The algorithm reads as follows:

**Algorithm 4:** Projected gradient descent to solve the projection problem.

**Input**:

An initial parameterization $q \in \mathcal{Q}_p$

A number of iterations $n_{it}$.

**Output**:

An approximation $\tilde{q}$ of the solution $q^*$ of (6.12)

**for** $k = 1$ *to* $n_{it}$ **do**

$$q^{(k+1)} \in \Pi_{\mathcal{Q}_p}\left(q^{(k)} - \tau \nabla (J_1 - J_2)(q^{(k)})\right) \tag{6.13}$$

The step-size $\tau$ should be selected depending on the regularity of the kernel $h$. Note that $\mathcal{Q}_p$ has no reason to be convex in general and the projection on $\mathcal{Q}_p$ might therefore not be unique. This explains the sign $\in$ instead of $=$ in Equation (6.13). If $\tau$ is well chosen, this algorithm is shown to converge to critical points of (6.12) in (Chauffert et al., 2015a). Let us finally mention that computing $\nabla J_1$ and $\nabla J_2$ is also a complicated issue that requires the use of tools developed for particle simulations. In this work, we use the parallelized non-uniform fast Fourier transform (Keiner et al., 2009; Teuber et al., 2011).

### 6.5.3 Discretization of the parameterization sets

In this section, we explicitly give the expressions of $\mathcal{Q}_p$ and $\Pi_{\mathcal{Q}_p}$ for the measures sets given in paragraph 6.4.1.

**Isolated points.** In the context of isolated points, $\mathcal{Q}_p = \Omega^p$, hence the projection $\Pi_{\mathcal{Q}_p}$ is the identity on $\Omega^p$. The updating step in Algorithm 4 is then $q^{(k+1)} = q^{(k)} - \tau \nabla (J_1 - J_2)(q^{(k)})$.

**Admissible curves for MRI.** The projection in the case where $\mathcal{P} = \mathcal{S}_T$ is the topic of (Chauffert et al., 2014b). The discretization of an element of $\mathcal{S}_T$ is a vector of $\mathbb{R}^{p \cdot d}$ where $d$ is the space dimension and $p$ is the number of points. Let $s(i)$ denote the curve location at time $(i-1)\delta t$ with $\delta t = \dfrac{T}{p-1}$. We define the first-order derivative by:

$$\dot{s}(i) = \begin{cases} 0 & \text{if } i = 1, \\ (s(i) - s(i-1))/\delta t & \text{if } i \in \{2, \dots, p\}. \end{cases}$$

In the discrete setting, the first-order differential operator can be represented by a matrix $\dot{\mathbf{M}} \in \mathbb{R}^{p \cdot d \times p \cdot d}$, i.e. $\dot{s} = \dot{\mathbf{M}} s$. We define the discrete second-order differential operator by

$\ddot{\mathbf{M}} = -\dot{\mathbf{M}}^*\dot{\mathbf{M}} \in \mathbb{R}^{p\cdot d \times p\cdot d}$. The projection problem is:

$$\Pi_{\mathcal{Q}_p}(c) = \underset{\substack{\|\dot{\mathbf{M}}s\| \leqslant \alpha \\ \|\ddot{\mathbf{M}}s\| \leqslant \beta}}{\arg\min} \|s - c\|_2^2$$

This problem can be solved using an iterative algorithm, resorting to the dual formulation of the problem (Chauffert et al., 2014b).

**Lines of variable length.** In this case, the admissible set is the set of measures supported by $N$ segments of variable length. Assuming that each segment is discretized in $k$ points, the total number of discretization points is $p = kN$ and the set can be written as follows

$$\mathcal{Q}_p(\mathcal{L}^N) = \{q \in \Omega^{p\times d},\, q_j = q_i + \frac{j-i-1}{k-1}\left(q_{i+k} - q_i\right), \\ \text{for } i = 1:k:kN \text{ and } i \leq j \leq i+k\},$$

where $1:k:kN$ denotes the set $\{1, k+1, 2k+1, \ldots, (N-1)k+1\}$.

The projection onto this set can be explicitly computed, as presented in Algorithm 5. For the sake of clarity, Algorithm 5 presents the projection onto the set of measures supported only by one segment in two dimensions.

**Algorithm 5:** Projection on $\mathcal{Q}_p(\mathcal{L})$.

**Input**:

$u$ : a vector of $k$ points

**Output**:

$q$ : a vector of $\mathcal{Q}_p(\mathcal{L})$

- Compute $C = k(k^2 - 3k + 2)/(6(k-1)^2)$
- Compute $D = k(2k^2 - 3k + 1)/(6(k-1)^2)$
- Compute $x_i^{(1)} = (k-i)u_i$ for $1 \leq i \leq k$
- Compute $x_i^{(2)} = (i-1)u_i$ for $1 \leq i \leq k$
- Compute $s^{(1)} = \sum_{i=1}^{k} x_i^{(1)}$
- Compute $s^{(2)} = \sum_{i=1}^{k} x_i^{(2)}$
- Evaluation of the end points
  - $q_k = C/(C^2 - D^2)\left(s^{(1)} - D/Cs^{(2)}\right)$
  - $q_1 = 1/C(s^{(2)} - Dq_k)$
- Place $(q_i)_{2 \leq i \leq k-1}$ uniformly distributed on $[q_1, q_k]$

### 6.5.4 Implementation details

Solving the projection problem (6.9) is computationally intensive. Hopefully, the design of sampling patterns is performed offline and large computing times are therefore acceptable. In practice, we used a workstation with 192 Gb of RAM and 32 Cores at 2.4 GHz. All codes were multithreaded.

The computing times varied from 2 hours to generate the sampling schemes for low resolution images proposed in Figure 6.5 up to 48 hours for the schemes adapted to very high resolutions images in Figure 6.12. In practice, we used 4,000 iterations to generate the sampling schemes with isolated measurements. For the sampling schemes composed of lines or curves, we used a multi-resolution strategy: we first optimize a undersampled curve and progressively oversampled it, thus reducing the number of iterations as the resolution increases. We observed that this strategy provides improved results and speeds up convergence.

## 6.6 Results

In this section, we test the proposed ideas for reconstructing a 2D image (i.e. a slice) of a brain phantom at 2 different resolutions on a field of view of 20 cm. In all experiments, we used the analytical phantoms provided in (Guerquin-Kern et al., 2012).

The first image is of size $256 \times 256$, which corresponds to an isotropic resolution of $780 \times 780\,\mu$m. This is a rather high but still feasible resolution for actual MRI scanners. The second image size is $2048 \times 2048$, which corresponds to an isotropic resolution of $98 \times 98\,\mu$m. The latter is really uncommon in the literature and is actually an important challenge since it might allow us to uncover the meso-scale of brain organization. For instance, (Fatterpekar et al., 2002) reported ex-vivo experiments on brains at a resolution of $78 \times 78 \times 500\,\mu m$ allowing to much better understand the cytoarchitecture of the human cortex. However, such spatial resolution cannot be achieved during in-vivo experiments owing to the very long scanning times. For instance, the images used in (Fatterpekar et al., 2002) took more than 14 scanning hours. Compressed sensing may therefore play a key role in the future to push forward such resolutions, especially with the emergence of ultra-high field MRI at 7T or even 11.7T in the near future. Moreover, recent theoretical results (Roman et al., 2014) suggest that compressed sensing should be used as a resolution enhancer rather than a time saver.

### 6.6.1 Constraints used in our experiments

To apply our projection algorithm, the kinematic constraints have to be specified. To this end and in order to make our numerical experiments realistic, we used typical constraints met on real MRI scanners. The kinematic constraints imposed by MRI acquisition are the gradient magnitude and slew-rate: here, we set $G_{\max} = 40\,\text{mT.m}^{-1}$ and $S_{\max} = 150\,\text{mT.m}^{-1}\text{.ms}^{-1}$. For proton imaging, $\gamma = 42.576\,\text{MHz.T}^{-1}$, which allows to compute $\alpha = \gamma G_{\max}$ and $\beta = \gamma S_{\max}$ in Equation (6.2). In addition to those constraints, we imposed our trajectories to last less than $200\,\mu$s[2] to keep a significant amount of signal. This constraint is particularly relevant in the context of echo planar imaging (EPI) where a full 2D $k$-space plane is sampled from a single radio frequency pulse. Indeed, in other classical acquisition scenarios (line-by-line sampling), line-dependent radio-frequency pulses are delivered which means that a single trajectory does not cover the complete 2D $k$-space but instead a single line only.

[2]Beyond this limit, the $T_2^*$ relaxation decay makes the noise predominant.

### 6.6.2 Empirical choice of the target density $\pi$

The theorems in Section 6.3 provide some general guidelines to design a reasonable density. However, finding the *best* target density $\pi$ is still an open issue depending on the number of measurements, the sparsity basis and the signal's structure.

In this paper, we therefore used an empirical method. The basic idea was to optimize $\pi$ experimentally in the family of polynomially decaying densities of type $1/(|k|+1)^\eta$. Those simple parametric densities have been used a lot in recent articles and have proved their efficiency in practice. Note however that they increase fastly at the origin, leading to high samples concentrations. Even though there is no formal proof of this fact, we observed that such concentrations are deleterious. The basic reason is that they bring more information than necessary for low frequencies, which in turn, reduces the number of samples available for higher frequencies.

Given an initial discrete distribution $\boldsymbol{\pi}_\eta$ with a profile proportional to $1/(|k|+1)^\eta$, we therefore constructed a *truncated* version $\tilde{\boldsymbol{\pi}}_\eta$ of $\boldsymbol{\pi}_\eta$ defined by

$$\tilde{\boldsymbol{\pi}}_\eta = \min(\lambda \boldsymbol{\pi}_\eta, \tau) \tag{6.14}$$

where $\lambda$ is chosen in such a way that $\|\tilde{\boldsymbol{\pi}}_\eta\|_1 = 1$. The distribution $\tilde{\boldsymbol{\pi}}_\eta$ has all components less than $\tau$, and approximates $\boldsymbol{\pi}_\eta$.

In all our experiments, the threshold $\tau$ is chosen in such a way that the expectation of the number of samples in each pixel does not exceed 4 for an i.i.d. drawing. Assuming that $\boldsymbol{\pi} \in \mathbb{R}^n$ where $n$ is the number of pixels in the image, it means that $m\tau = 4$, where $m$ is the number of drawn samples. An illustration of density (6.14) is given in Figure 6.3.

### 6.6.3 New sampling patterns

We designed sampling schemes with the proposed algorithm and compared them to the state-of-the-art on the reconstructed brain phantom images. We compared 6 sampling patterns identified by letters:

- Standard patterns:
  - (a) **Independent and identically distributed drawings** according to a prescribed density $\pi$. This is the pattern considered in most compressed sensing theories. This pattern is not feasible in reasonable times, but serves as a reference.

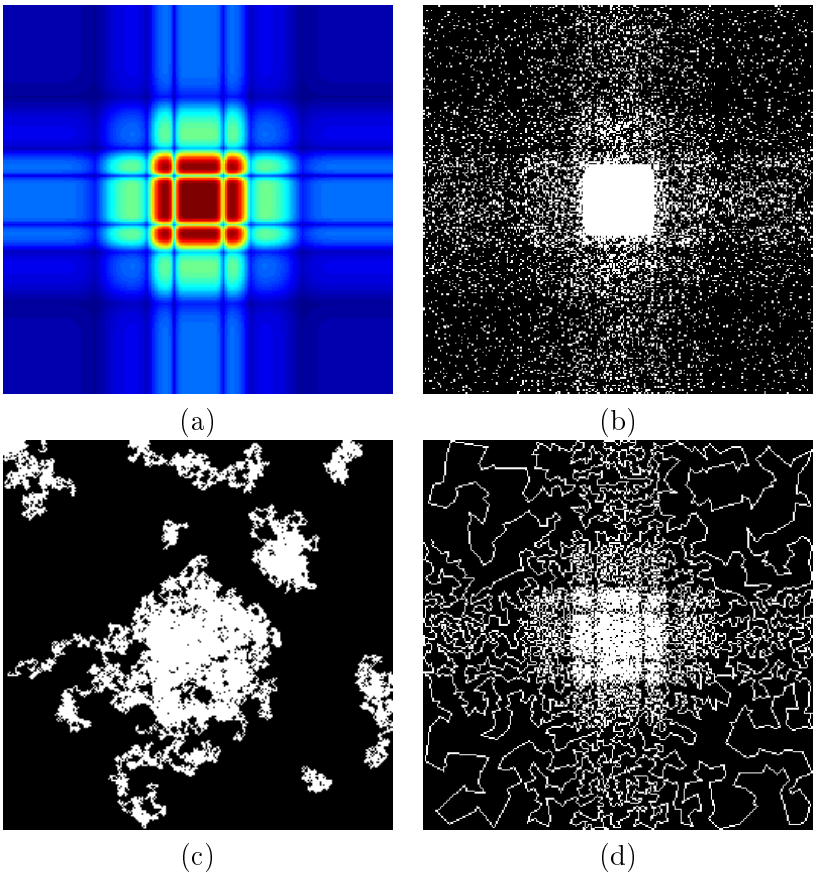


FIGURE 6.2: A few variable density samplers. (a) density $\pi$. (b) $\pi$-variable density sampler with i.i.d. drawings. (c) $\pi$-variable density sampler with a Markov chain. (d) $\pi$-variable density sampler with a traveling salesman problem solution.

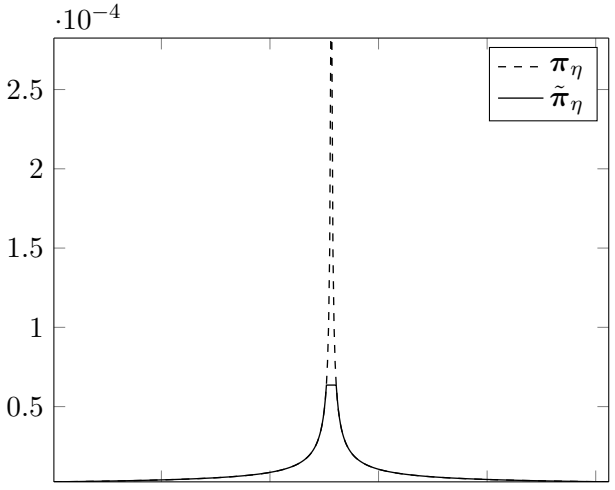


FIGURE 6.3: Action of the thresholding algorithm. The initial density $\boldsymbol{\pi}_\eta$ in dashed line and its thresholded version $\tilde{\boldsymbol{\pi}}_\eta$ defined in (6.14) in solid line.

  - (b) **Equispaced radial lines**. This is another sampling pattern that is commonly used in MRI (Winkelmann et al., 2007). We consider that a spoke is a line composed of $n/2$ samples, such that the distance between two samples is the pixel size.
  - (c) **Spiral sampling**. We consider a spiral with the chosen target density $\pi$ (see (Chauffert et al., 2014b)), and reparameterize it to be admissible (Lustig et al., 2008). We replicate and rotate it a few times, to obtain a pattern made of interleaved spirals.

- Measure projection patterns:
  - (d) **Projection of $\pi$ on the set of isolated measurements** defined in (6.7).
  - (e) **Projection of $\pi$ on the set of lines with varying lengths**. It is denoted $\mathcal{M}(\mathcal{L}^N)$ and defined in (6.8). Each line contains the same number of samples $n/2$ as a radial spoke.
  - (f) **Projection of $\pi$ on the set of admissible curves $\mathcal{S}_T$**, defined in Equation (6.2).

#### 6.6.3.1 Standard resolution imaging

In this section, we focus on the reconstruction of $256 \times 256$ images. We imposed each trajectory to contain exactly 8192 $k$-space points, which corresponds to a typical value of the maximal buffer size. We also fixed the sampling rate to $\Delta t = 20\,\mu$s. In practice, this would ensure a high signal-to-noise (SNR) ratio. The trajectories length was therefore 164 ms. With this choice of $\Delta t$ and the number of pixels for a given slice, a full acquisition of the Cartesian $k$-space would take around 1.3 s using an EPI trajectory (the fastest full $k$-space acquisition scheme). Since the MR signal is not available for that period given the $T_2^*$ decay, this means that at this spatial resolution, EPI is not feasible as such and require a segmented acquisition strategy with 8 segments. The sampling patterns introduced hereafter contain $m = 16384$ samples, i.e. 25% of the size of the Cartesian grid. This also corresponds to 2 curves of 8192 samples, i.e. an acquisition of about 328 ms (i.e. 4 times faster than the EPI scenario). For this resolution, we found out that the best decay $\eta$ defined in Section 6.6.2 is $\eta = 1.5$.

**Description of experiments.** We compared the different sampling patterns described above. To achieve a fair comparison, the number of samples (16384) and the sampling rate ($20\,\mu$s) were fixed in each experiment wherever it was meaningful (i.e. except for

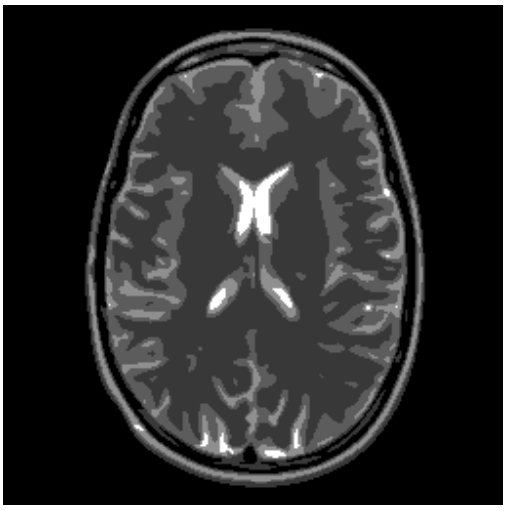

FIGURE 6.4: Axial slice of the phantom used in the experiments of size $256 \times 256$. The left brain hemisphere is shown on the right: left is right.

i.i.d. or isolated points). For the spiral experiment, we considered two interleaved spirals traversed in 164 ms each.

Data were simulated using the phantom depicted in Figure 6.4. The inverse problem used to reconstruct an image from simulated $k$-space data is Problem (6.4). Parameter $\lambda$ parameter was selected by hand once for all so as to nearly reach the equality constraint $\mathbf{S\Psi x} = \mathbf{y}$ and to provide a visually satisfactory solution in less than 1,000 iterations.

**Results.** In Figure 6.6, we show the reconstruction results for the different sampling schemes depicted in Figure 6.5. Let us summarize the main conclusions.

- First, we noticed that the two schemes composed of isolated measurements provided rather satisfactory reconstruction results despite a few artifacts (17.7 and 18.3 dB in (a) and (d), respectively) with one fourth of the measurements. This is an appealing result, but unfortunately the schemes cannot be implemented on a scanner.
- The repulsion between isolated samples in (d) improved the reconstruction result slightly by 0.6 dB. This result tends to validate the interest for this strategy and for a good coverage of the sampling space.
- Classical sampling patterns were feasible and reduced the scanning times by a factor 4, but provided results that may not be considered as sufficient by clinicians (15.4 dB for radial in (b) and 13.2 dB for spirals in (c)). The reconstruction based on radial lines induced many small artifacts while the reconstruction based on spirals suffered from ringing.

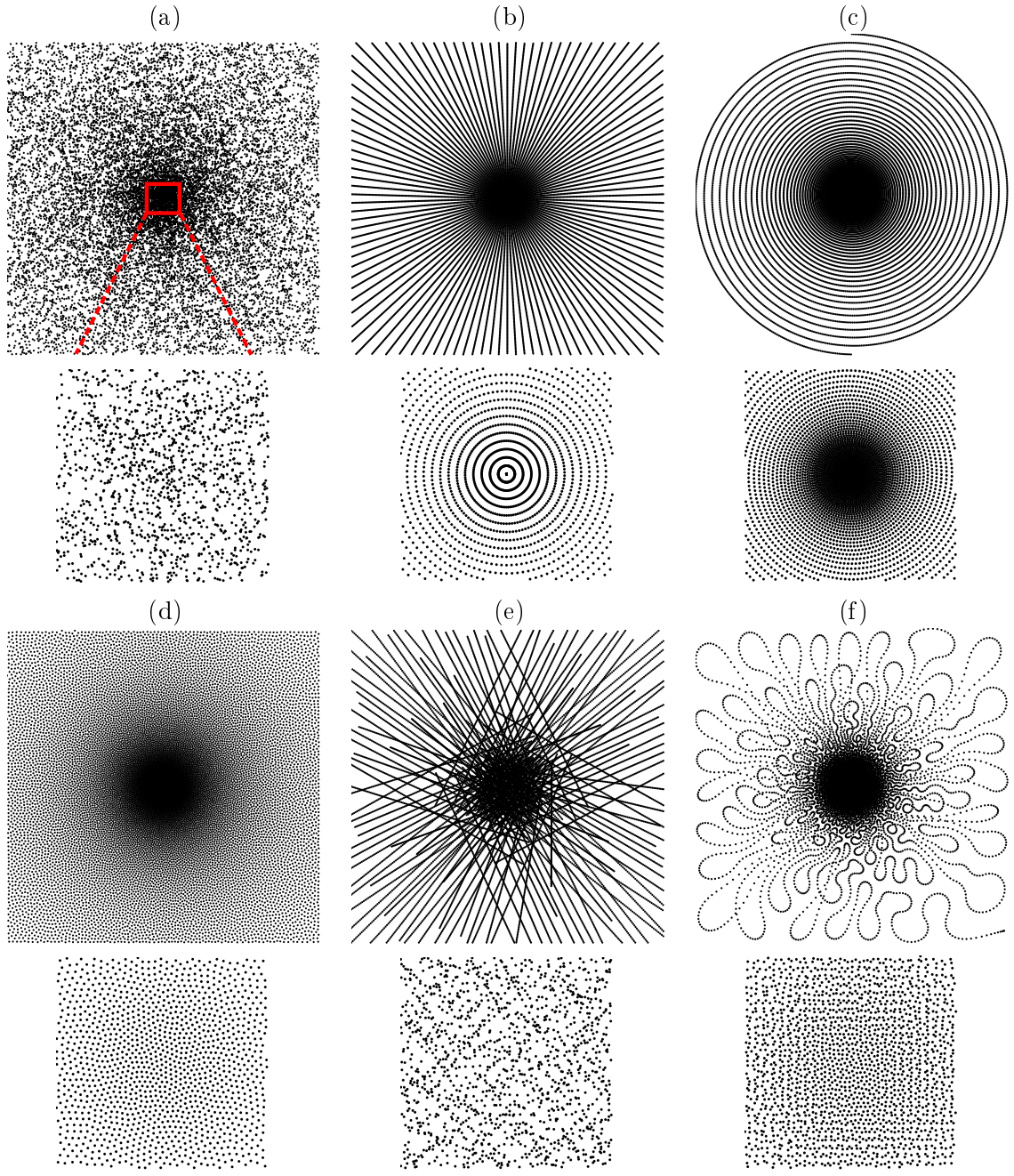


FIGURE 6.5: Classical sampling schemes (a-c) and sampling schemes obtained with the proposed projection algorithm (d-f). Top row: (a): independent drawing; (b): radial lines ; (c): spiral trajectory. Second row: zooms on the $k$-space centers. Third row: (d): isolated points; (e): isolated lines of fixed length; (f): admissible curves for MRI. Bottom row: zooms on the $k$-space center. Corresponding reconstruction results are provided in Figure 6.6.

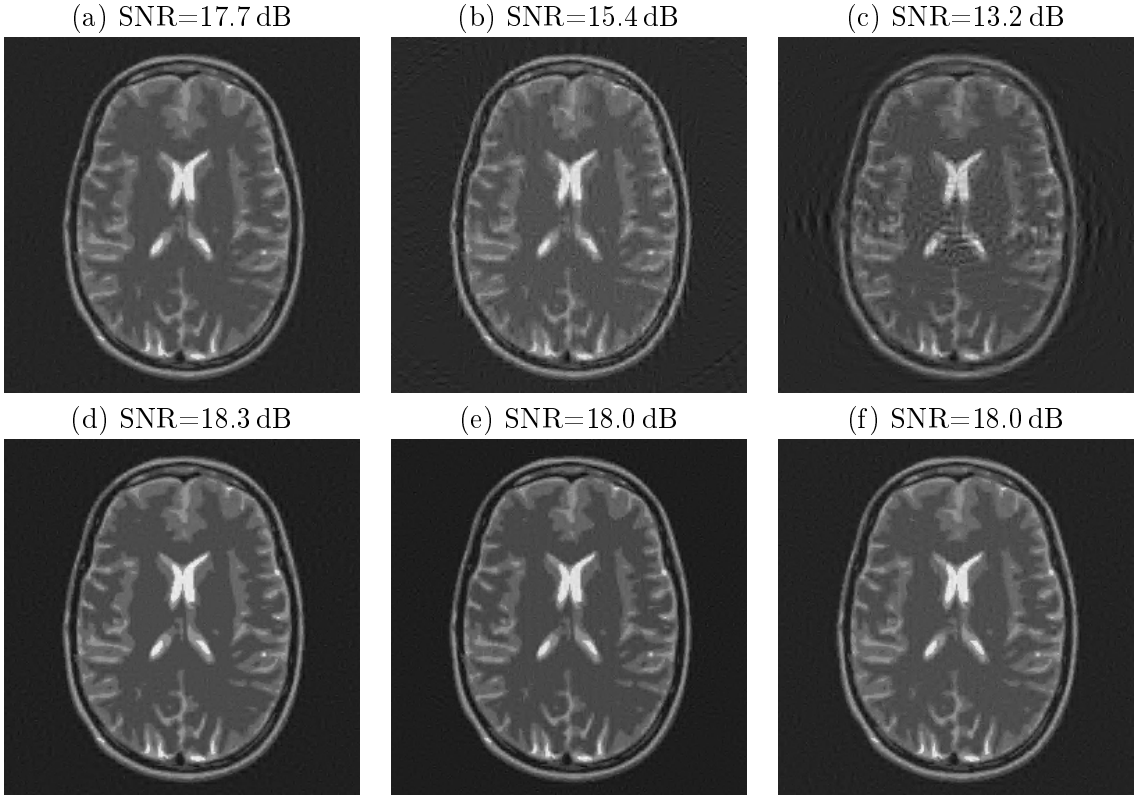


FIGURE 6.6: Reconstruction results for the sampling patterns proposed in Figure 6.5 on the phantom (Figure 6.4). Left is right.

- Quite unexpectedly, the new sampling patterns generated by our algorithm yielded improved reconstruction results as compared to i.i.d. drawings. The latter sampling scheme is often considered as the best existing undersampling strategy. This result shows that adding complicated but realistic sampling constraints can still permit to get competitive reconstruction results. In particular, the sampling pattern in Figure 6.6(f) took only one fourth of the scanning time and yielded satisfactory reconstructions.

#### 6.6.3.2 Very high resolution imaging

Here, we focussed on the reconstruction of very high resolution ($2048 \times 2048$) images. We used the same constraints as before including the maximum duration of 200 ms per trajectory. We decreased the sampling period down to $\Delta t = 8\,\mu$s which corresponds to the minimal value of a clinical scanner. We no longer managed the buffer size constraint as done in the previous section and performed experiments with 100,000 and 200,000 measurements. This corresponds to 2.4% and 4.8% of the total number of pixels in the image respectively. This also corresponds to a total sampling duration of 0.8 s or 1.6 s,

which again might be feasible on a real scanner considering a segmented acquisition scheme, with 4 and 8 segments, respectively.

The radial lines were composed of 1024 equispaced measurements. The distance between two measurements is 1 pixel. The spiral pattern was composed of 4 (resp. 8) rotated versions of an initial spiral designed as described at the beginning of Section 6.6.3 for 100,000 (resp. 200,000) measurements. The line of varying lengths were all composed of 512 samples. This corresponds to 196 lines for the 100,000 measurement experiments and 391 lines for the 200,000 experiment.

We aimed at reconstructing the phantom in (Guerquin-Kern et al., 2012). We modified it slightly by superimposing the high resolution text `COGITO ERGO SUM` to white matter in the left frontal region (see Figure 6.7).

Similarly to the previous section, the sampling density was optimized experimentally in the family of truncated, polynomially decaying densities of type $1/(|k|+1)^\eta$. For this resolution, the best decay was $\eta = 2$. The resulting patterns are shown at different resolutions in Figures 6.8–6.9 for 100,000 measurements and Figures 6.11–6.12 for 200,000 measurements. For each scheme we reconstructed a $2048 \times 2048$ image by solving Problem (6.4).

Let us summarize our main observations.

- The use of 200,000 measurements yielded significantly better reconstruction results than 100,000 samples. However, the relative differences between the sampling schemes did not vary between the two sampling ratios. In what follows, we therefore draw conclusions that are valid for both.
- Similarly to the standard resolution experiment, sampling schemes made of i.i.d. drawings significantly outperformed radial lines and spirals sampling.
- Radial lines performed particularily poorly. This was probably due to the fact that for this resolution, the best sampling decay was $\eta = 2$, whereas we found $\eta = 1.5$ for the standard resolution experiment. Note that radial lines have a slow decay of order $1/|k|$, which might explain the observed discrepancy. Also note that the embedded text for radial reconstruction was readable, whereas it was not for spiral sampling. Once again, this is very likely a consequence of the slower decay for the sampling density. On the contrary, the cortex is not correctly recovered by radial lines, while the reconstruction is acceptable for spirals. This experiment thus suggests that the sampling density should depend on the relative importance of low and high resolution details.

- The repulsed isolated measurements scheme performed slightly better than i.i.d. drawings, but not significantly so.
- Similarly to the previous section, the sampling schemes generated by our algorithm performed significantly better than spiral and radial patterns. The gain ranged from 1.7 dB to 3.6 dB which is significant, since they require the same scanning time.
- In contrast to the previous section, we observed that the feasible sampling schemes performed significantly worse than i.i.d. drawings in terms of SNR. A reason that might explain this behavior was that $\Delta t = 8\,\mu$s for this resolution while we used $\Delta t = 20\,\mu$s in the previous experiment. This means that the distance between consecutive samples was more than twice smaller (harder constraint). It is also important to realize that, although the differences between reconstructions were strong in terms of SNR, the visual perceptual differences mainly rely on small artifacts which do not severely degrade image analysis.
- The results obtained with 200,000 samples were of a high quality, despite the realistic sampling constraints added. This very positive result suggest that obtaining $2048 \times 2048$ images might be feasible in 1.6 s obly using a segmented acquisition (8 segments) scheme. This should be definitely deemed as a major advance for MRI. Of course, these results were preliminary since we did not manage all degradations appearing on actual scanners such as noise, Eddy currents ...
- Finally, it is possible to infer the gain in terms of scanning times using the proposed approach by comparing Figure 6.10 and Figure 6.13. The SNR of the reconstructed image with 4 admissible curves and 0.8 seconds is 20.7 dB (see Figure 6.10 (f)). To reach the same quality, radial lines and spirals need roughly twice as long (see Figure 6.13 (b) and (c)). This result shows that the proposed ideas may reduce the actual scanning times by a factor 2 compared to existing approaches.

## 6.7 Conclusion

This paper has provided an overview of existing compressed sensing results for MRI, both from theoretical and practical points of view. We also proposed an original approach to design efficient sampling schemes complying with the physical constraints of MRI scanners. Even though we focussed on standard anatomical MRI, the proposed ideas could be used, with some adaptations, in nearly all MRI fields (functional imaging, diffusion-weigted imaging, perfusion imaging, ...) and might have applications well beyond.

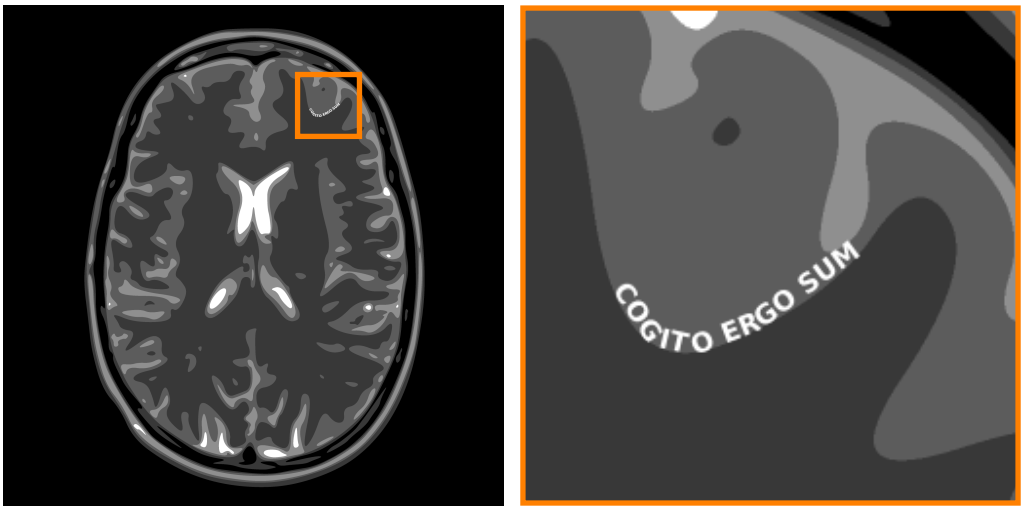


FIGURE 6.7: Axial slice of the brain phantom used in our 2048×2048 images (left) with a magnification on the left frontal area where the text has been superimposed (right).

.

The numerical procedure we proposed for generating sampling schemes was based on a projection of sampling distributions onto a set of admissible measures using a taylored dissimilarity measure. Even though computationally intensive, this algorithm was able to solve very large scale problems and could be extended to 3D quite easily. Probably the most promising result of this paper is practical: we showed through simulations that 1.6 s using a multi-shot acquisition (8 segments) might be enough to reconstruct a very high resolution slice of size $2048 \times 2048$. The validity of this result will be tested quite soon on the 7T scanner of NeuroSpin to check whether this constitutes a major improvement over existing sampling strategies which currently need a dozen of hours to reconstruct a hundred slices at this spatial resolution.

## Acknowledgments

The authors wish to thank Daniel Potts, Toni Volkmer, Gabriele Steidl for their support and help to run the excellent NFFT library (Keiner et al., 2009). They thank Anders Hansen for his review and insights on a preliminary version of this paper. They thank Alexandre Vigneau for his numerous comments on the physics of MRI scanners.

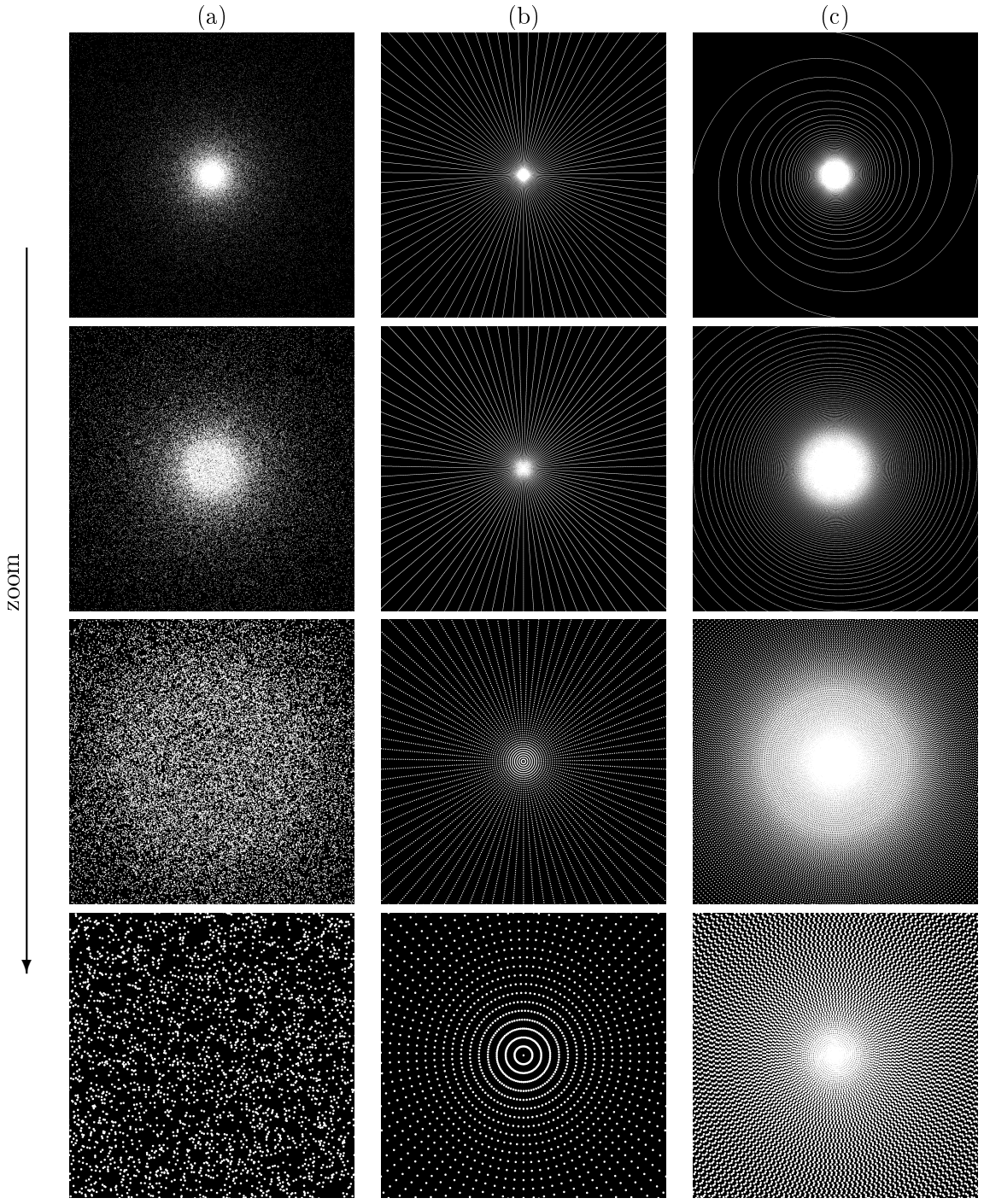


FIGURE 6.8: Standard sampling schemes composed of 100,000 samples. (a): i.i.d. drawings. (b): Radial lines. (c): 4 interleaved spirals.

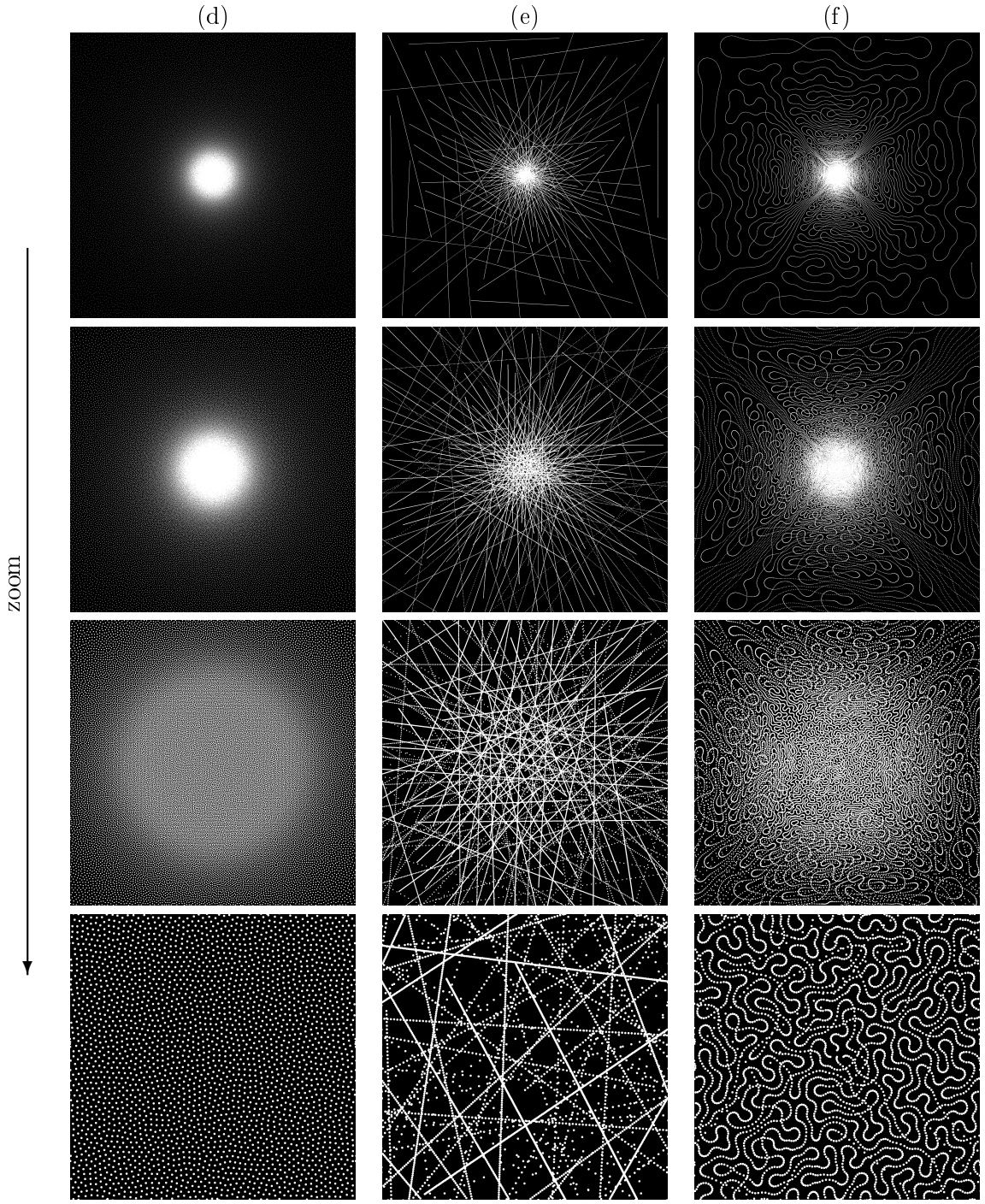


FIGURE 6.9: Sampling schemes yielded by our algorithm and composed of 100,000 samples. (d): Isolated measurements. (e): Lines of variable length. (f): 4 feasible curves in MRI.

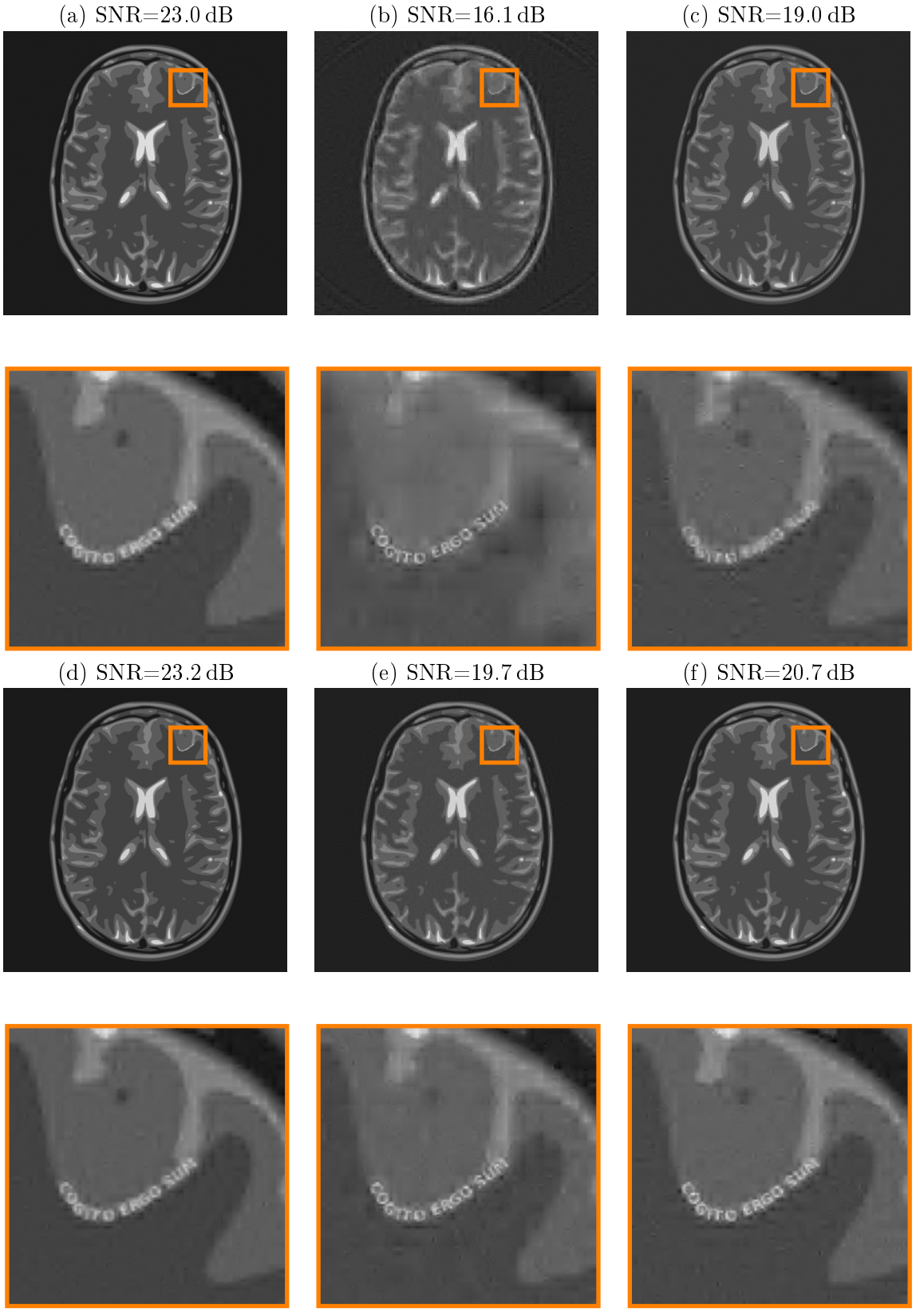


FIGURE 6.10: Very high resolution reconstructions using 100,000 samples (2.4% of the number of pixels) and different sampling schemes. Letters correspond to Figures 6.8–6.9.

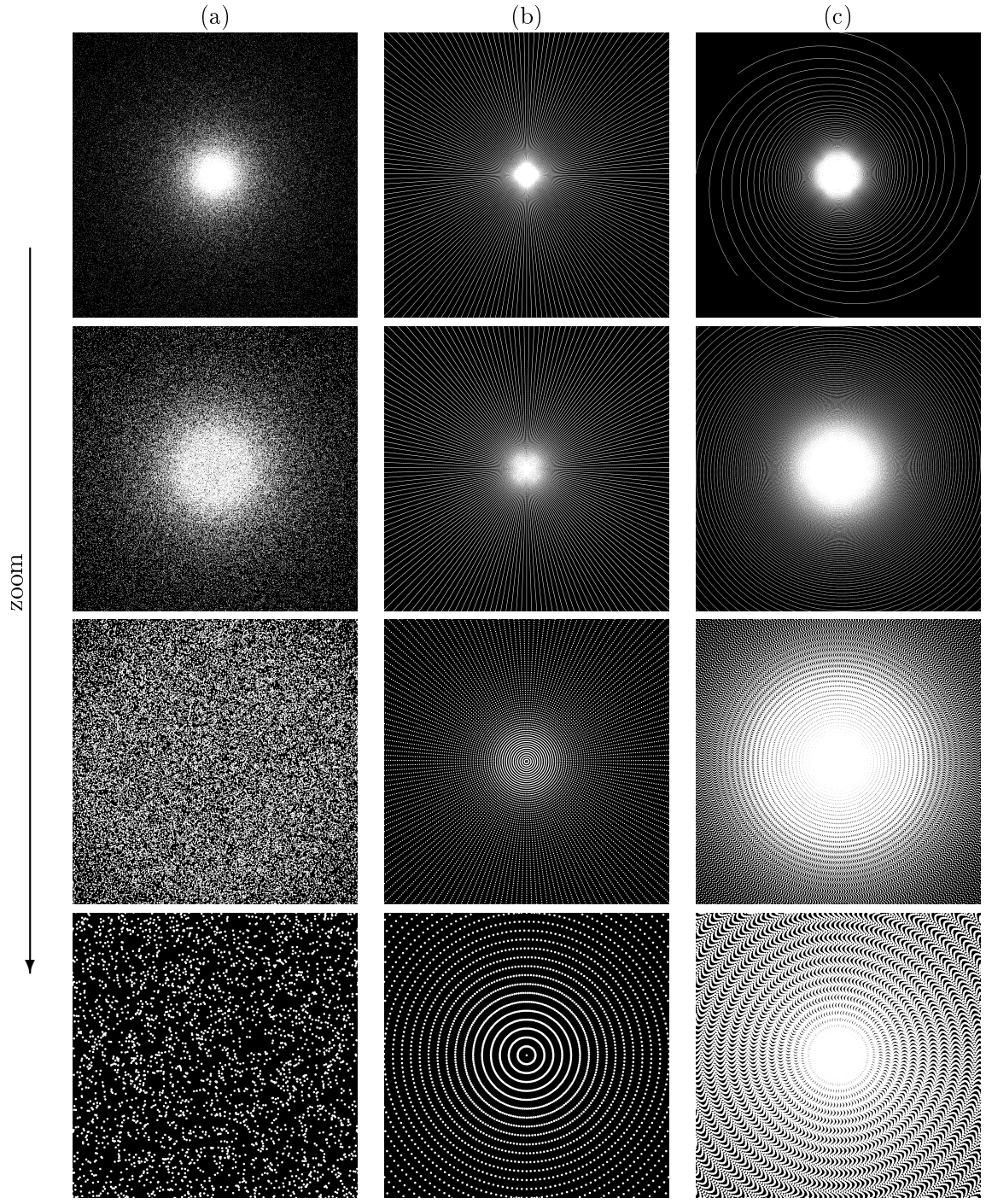


FIGURE 6.11: Standard sampling schemes composed of 200,000 samples. (a): i.i.d. drawings. (b): Radial lines. (c): 8 interleaved spirals.

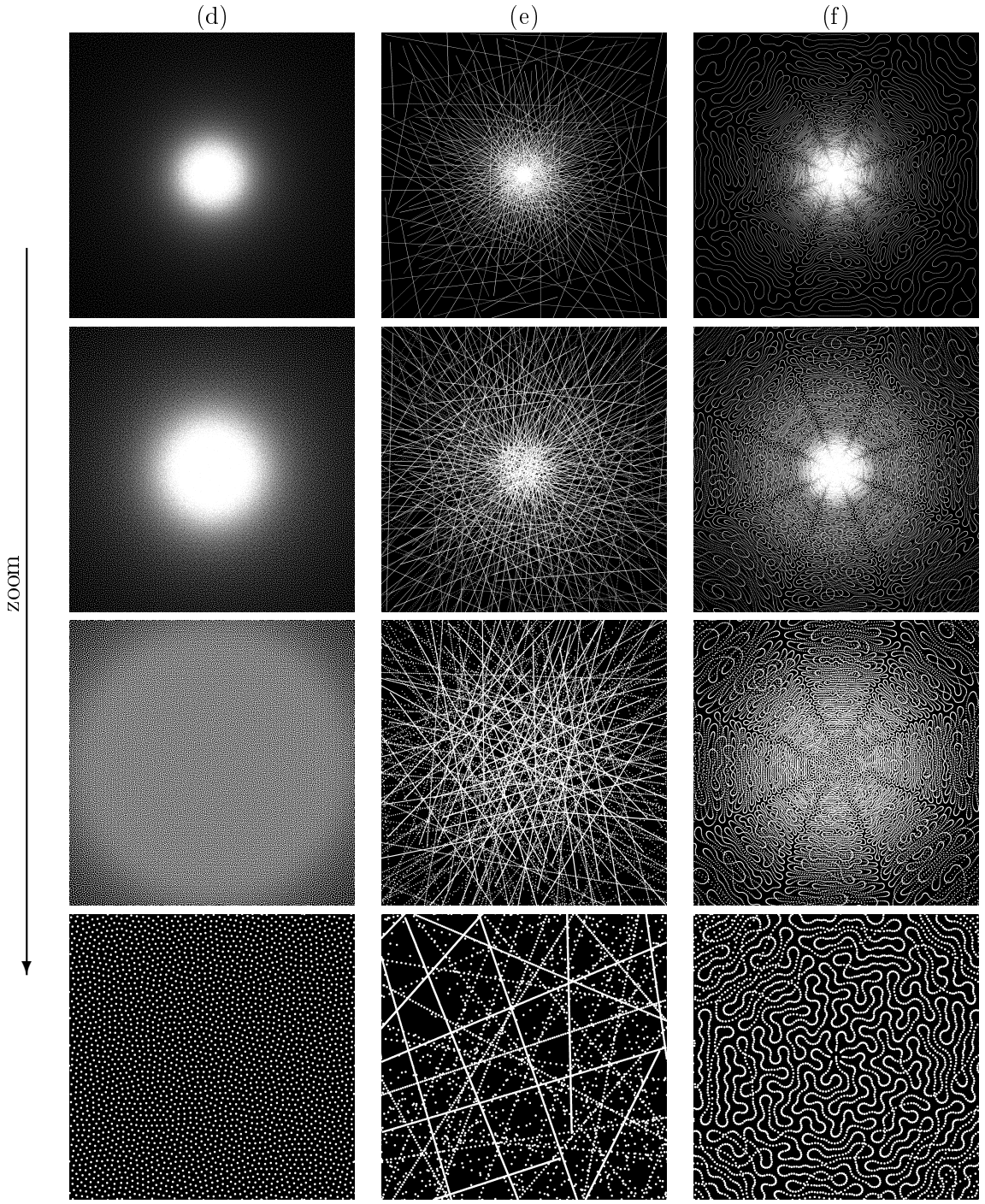


FIGURE 6.12: Sampling schemes yielded by our algorithm and composed of 200,000 samples. (d): Isolated measurements. (e): Lines of variable length. (f): 8 feasible curves in MRI.

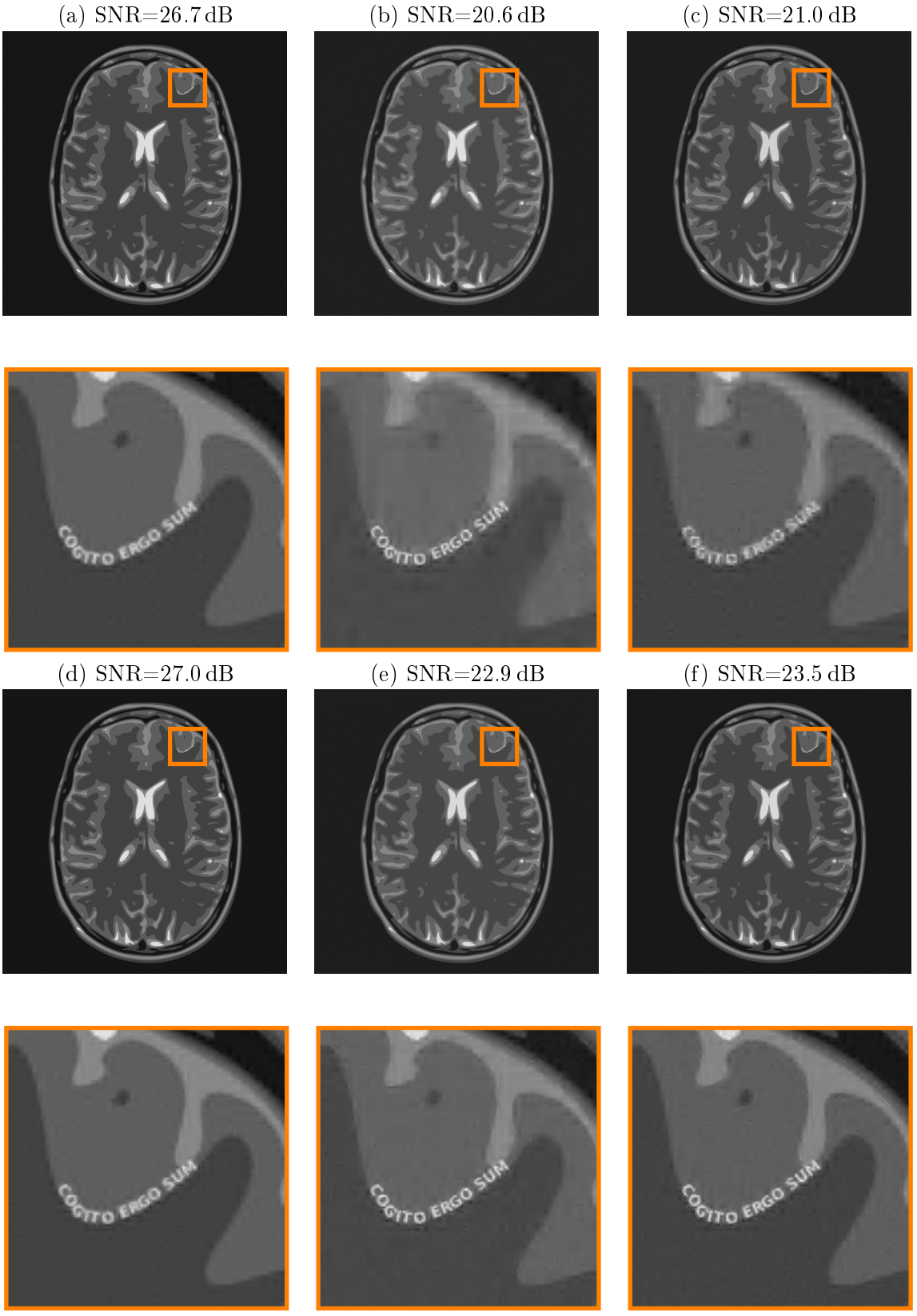


FIGURE 6.13: Very high resolution reconstructions using 200,000 samples (4.8% of the pixels number) and different sampling schemes. Letters correspond to Figures 6.11–6.12.

# Chapter 7

# Résumé des chapitres

## Résumé du chapitre 2

Ce chapitre se repose sur l'article (Chauffert et al., 2014a). L'objectif est ici de montrer que l'application directe de l'échantillonnage compressif n'est pas possible en imagerie par résonance magnétique. En effet, les bases de représentation et d'acquisition ne sont pas incohérentes d'une part, et les schémas obtenus à partir de tirages i.i.d. comme décrits par la théorie de l'échantillonnage compressif ne sont pas physiquement admissibles en IRM d'autre part. Dans cet article, nous justifions que l'échantillonnage déterministe du centre de l'espace-$k$ permet de répondre à la première question, et nous donnons une expression de la distribution *optimale* $\pi$ qui permet de réduire le nombre de mesures nécessaires à la reconstruction exacte. Cette distribution, représentée Fig 7.1, vérifie $\pi_i \propto \|a_i\|_\infty^2$, où $a_i$ est la $i$-ème ligne de la matrice $A = F^*\Psi$. D'autre part, nous définissons la notion d'échantillonneur à densité variable. Cette notion s'adapte également aux trajectoires continues qui sont une condition nécessaire (mais non suffisante) pour vérifier les contraintes cinématiques sur la trajectoire imposés par les gradients de la machine. Nous donnons deux stratégies pour générer des schémas continus à densité variable. L'une est basée sur des marches aléatoires et l'autre sur une solution de voyageur de

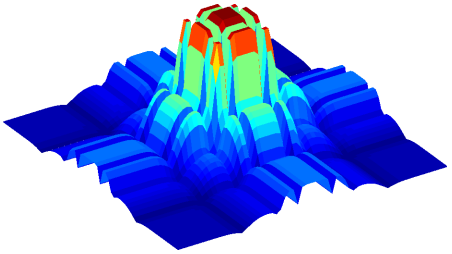

FIGURE 7.1: Distribution optimale $\pi$ pour la matrice $A = F^*\Psi$.

commerce. Pour la première, nous donnons un résultat d'échantillonnage compressif qui fait dépendre le nombre de mesures en $1/\epsilon(\mathbf{P})$ où $\epsilon(\mathbf{P})$ est le trou spectral de la chaîne de Markov associé à la marche aléatoire. Le résultat présenté ici peut être amélioré, cf. annexe 3. Des exemples de chaînes de Markov sont donnés dans la Fig. 7.2. Dans un second temps, nous relions la distribution de tirage des « villes » $p$ du voyageur de commerce à celle de la courbe $q$ par la relation $q \propto p^{\frac{d-1}{d}}$ où $d$ est la dimension. En particulier, en 2D, il faut tirer les villes selon la distribution proportionnelle à $\pi^2$ pour obtenir un échantillonnage à la densité $\pi$, comme illustré Fig. 7.3(a,b).

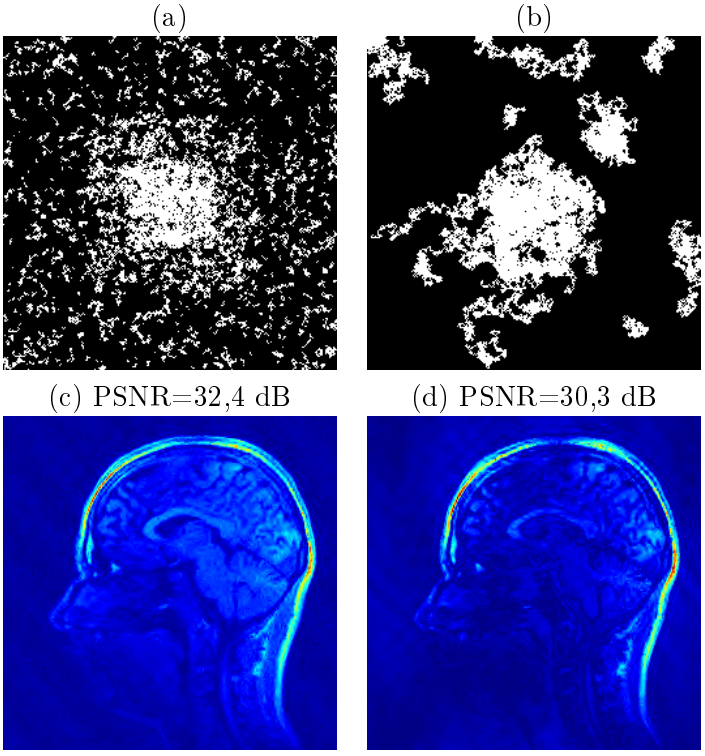


FIGURE 7.2: Schémas d'échantillonnage reposant sur des chaînes de Markov de longueur moyenne 10 (a) ou 1000 (b), et reconstruction d'images après échantillonnage selon les schémas (a) et (b).

Nous décrivons deux quantités centrales d'un processus à densité variable, à savoir :

- Sa densité cible. Nous montrons théoriquement et par des expériences numériques que la densité doit décroître des basses vers les hautes fréquences. En effet, les vecteurs de la transformée en ondelettes et de la transformée de Fourier sont plus cohérents aux basses fréquences (i.e., correspondent à des valeurs de $\|a_i\|_\infty^2$ plus grandes), et sont donc cruciaux car eux seuls contiennent les informations des coefficients d'ondelettes d'approximation. D'autre part, la plupart des l'information de

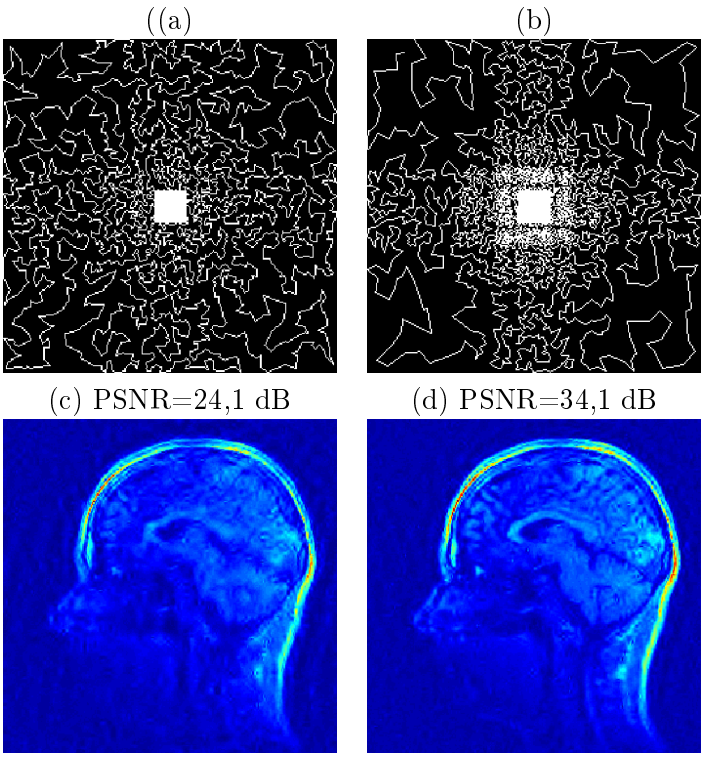


FIGURE 7.3: Exemples de schémas d'échantillonnage reposant sur une solution du voyageur de commerce dont les villes sont tirées selon $\pi$ (a) et $\propto \pi^2$ (b), et reconstruction des images après échantillonnage de l'espace-$k$ selon les schémas (a) et (b).

l'image étant contenue dans les basses fréquences, ceci renforce l'intérêt d'échantillonner plus les basses fréquences. L'importance de la densité d'échantillonnage sur les résultats de reconstruction est illustrée Fig. 7.3(c,d).

- Le temps de mélange. Cette grandeur mesure la vitesse de convergence de la mesure empirique de la courbe vers la mesure cible. Dans le cas des marches aléatoires, cette quantité vaut $1/\epsilon(\mathbf{P})$. Celle-ci représente également la rapidité pour le processus à recouvrir l'espace. Le cas optimal est un tirage i.i.d. qui correspond à $\epsilon(\mathbf{P}) = 1$. Le succès des solutions du voyageur de commerce s'explique par leur rapidité à couvrir rapidement l'espace, alors que les marches aléatoires laissent de grandes zones de l'espace-$k$ non explorées, ce qui explique les mauvais résultats en reconstruction. Ceci est illustré dans la Fig. 7.2, où plus les chaînes sont courtes, et plus le trou spectral est grand.

## Résumé du chapitre 3

Dans ce chapitre, nous montrons une nouvelle inégalité de concentration pour des chaînes de Markov à valeurs dans les matrices Hermitiennes. Considérons $\mathcal{G}$ un graphe à $N$

nœuds, et $X_1, \dots X_n$ une chaîne de Markov définie sur ce graphe, dont l'évolution est donnée par $X_1 \sim q$ et $P$ est la matrice de transition. Soit $\pi$ le vecteur de $\mathbb{R}^N$ tel que $\pi^T P = \pi^T$ ($\pi$ est appelée distribution stationnaire de la chaîne). Supposons enfin que la chaîne est réversible, c'est-à-dire que $\pi_i P_{ij} = \pi_j P_{ji}$.

Soit $f$ une fonction de $\mathcal{G}$ à valeur dans $\mathbb{H}^d := \{M \in \mathbb{C}^{d\times d}, M^* = M\}$. Supposons sans perte de généralité que $\sum_{y\in\mathcal{G}} \pi(y) f(y) = 0$, et notons $R = \sup_{y\in\mathcal{G}} \lambda_{\max}(f(y))$. Alors :

$$\mathbb{P}\left(\lambda_{\max}\left(\sum_{i=1}^{n} f(X_i)\right) \geqslant t\right) \leqslant d \cdot \sup\left(\frac{q_i}{\pi_i}\right) \cdot \exp\left(-\frac{\varepsilon t^2}{4\sigma_n^2 + 2Rt\varepsilon/3}\right). \tag{7.1}$$

où $\varepsilon$ est le trou spectral de la chaîne, i.e., la différence de ses deux plus grandes valeurs propres, et

$$\sigma_n^2 := n \cdot \lambda_{\max}\left(\sum_{y\in\mathcal{G}} \pi(y) f(y)^2\right).$$

Cette inégalité est ensuite utilisée pour démontrer un résultat d'échantillonnage compressif. Dans le cas d'un échantillonnage le long d'une chaîne de Markov comme décrit dans le chapitre 2, le nombre de mesures nécessaires pour garantir la reconstruction est $\mathcal{O}\left(\frac{s}{\varepsilon} \cdot \sum_{j=1}^{n} \|a_j\|_\infty^2 \cdot \log(6n/\eta)\right)$, ie, est plus grand que dans le cas i.i.d. d'un rapport $1/\varepsilon$, car $\varepsilon \leqslant 1$. Cette borne est cohérente car dans le cas i.i.d., $\varepsilon = 1$.

## Résumé du chapitre 4

Les résultats de ce chapitre reposent sur la soumission (Chauffert et al., 2014b). Dans le chapitre 2, nous proposons des trajectoires d'échantillonnage continues. Les contraintes physiques de l'IRM imposent que les dérivées première et seconde de la trajectoire soient bornées, i.e., une courbe admissible doit appartenir à :

$$\mathcal{S}_T = \left\{s : [0,T] \to \mathbb{R}^3, \|\dot{s}(t)\| \leqslant \gamma \cdot G_{\max}, \|\ddot{s}(t)\| \leqslant \gamma \cdot S_{\max}, \forall t \in [0,T]\right\}.$$

Pour une trajectoire donnée, une première approche (Lustig et al., 2008; Hargreaves et al., 2004) consiste à trouver la reparamétrisation optimale de la trajectoire. Nous proposons ici une approche alternative qui consiste à projeter une courbe paramétrée

initiale sur l'espace des contrainte $\mathcal{S}_T$ :

$$s^* = \arg\min_{s \in \mathcal{S}_T} \|s - c\|_2^2 :$$

Une illustration de la solution de ce problème de projection est donné (Fig. 7.4)

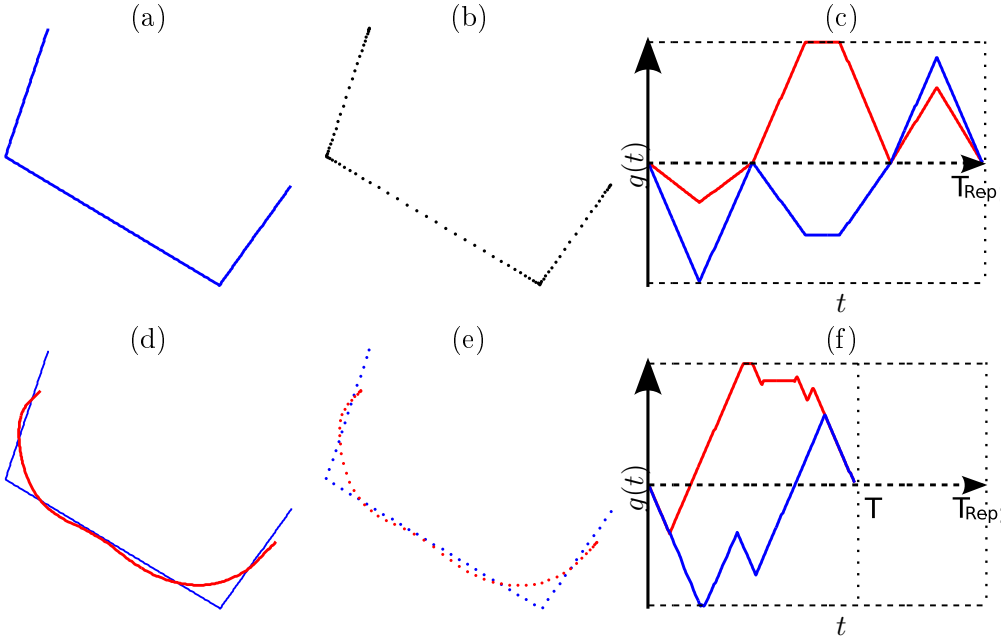


FIGURE 7.4: Exemple synthétique de la différence entre les deux approches : reparamétrisation optimale et projection. **Première ligne:** Reparamétrisation optimale (Lustig et al., 2008). (a): paramétrisation initiale. (b): représentation discrète de la reparamétrisation optimale de la courbe $\in \mathcal{S}_T$. (c): Gradients $(g_x, g_y)$. Les lignes pointillées correspondent à 0 and +/- $G_{\max}$. **Deuxième ligne:** Illustration de l'algorithme de projection. (d): support de la courbe d'entrée $c$ de l'algorithme paramétrisé à vitesse constante $\gamma \cdot G_{\max}$ et support de la projection $s^*$ sur $\mathcal{S}$. (e): représentation discrète de la courbe d'entrée et de la projection. (f): Gradients correspondants $(g_x, g_y)$ avec la même échelle temporelle qu'en (c): le temps de parcours de $s^*$ est plus court de *39%*.

Les avantages de l'approche par projection sont multiples :

- Le temps de parcours est fixé à l'avance dans cette méthode (c'est le même que celui de la paramétrisation initiale), alors qu'il dépend de la trajectoire pour la reparamétrisation optimale. En particulier, si la trajectoire comporte des parties à forte courbure (et dans un cas extrême, des points anguleux comme dans une trajectoire de voyageur de commerce ou EPI), la durée d'acquisition peut exploser et rendre cette méthode inexploitable.

- La distribution empirique de la courbe projetée est « proche » de celle de la courbe d'entrée.

- Cet algorithme de projection permet d'intégrer de nombreuses contraintes supplémentaires qui jouent un rôle important en IRM (cf chapitre 5).

Nous présentons des résultats de reconstruction en 2D sur différents types de trajectoires : EPI, spirale et solution du voyageur de commerce. Des exemples de reconstruction en 3D pour des trajectoires de voyageur de commerce projetées sur l'ensemble des contraintes, ainsi que des applications à l'angiographie sont également fournies.

La solution du problème de projection se calcule en écrivant le problème de projection dans l'espace dual où le problème a une structure favorable. Celui-ci s'écrit en effet comme somme d'un terme convexe différentiable et d'un terme convexe dont on sait calculer l'opérateur proximal.

## Résumé du chapitre 5

Dans ce chapitre, nous étudions un problème de projection de mesures. Soit $\Omega \subseteq \mathbb{R}^d$ le domaine de définition d'une mesure de probabilité $\pi$ et $h \in L^2(\Omega)$ un noyau. Le problème de projection s'écrit:

$$\inf_{\mu \in \mathcal{M}_N} \|h \star (\mu - \pi)\|_2 \tag{7.2}$$

où $\star$ est l'opérateur de convolution et $\mathcal{M}_N$ est un ensemble de mesures. Pour illustrer ces notions, nous considérons le cas où $h$ est un noyau gaussien, $\pi$ est une image (Fig. 7.5(a)) et

$$\mathcal{M}_N = \left\{ \frac{1}{N} \sum_{i=1}^{N} \delta_{p_i}, (p_i)_{1 \leqslant i \leqslant N} \in \Omega^N \right\}$$

où $\delta_{p_i}$ est une mesure de Dirac en $p_i$. Dans la Fig. 7.5(b), nous représentons la solution du problème de projection où les mesures de Dirac sont représentées par un disque noir. Les convolutions de $\pi$ et la solution sont représentées en Fig. 7.5(c) et (d).

La convolution par un noyau gaussien permet la comparaison entre les images (a) et (b) dont la différence est très grande. Le filtre gaussien est représentatif du système visuel humain et peut être exagéré en regardant les images (a) et (b) de plus loin et en plissant les yeux. Le problème de représenter une image par des points est connu sous le nom de *stippling* et est utilisé en imprimerie par exemple.

Dans ce chapitre, nous considérons des ensembles de mesures plus généraux que des sommes de mesures de Dirac. Définissons une paramétrisation comme une fonction $p : X \to \Omega$ et $\mathcal{P}$ un ensemble de paramétrisations. Alors nous pouvons définir $\mathcal{M}(\mathcal{P}) =$

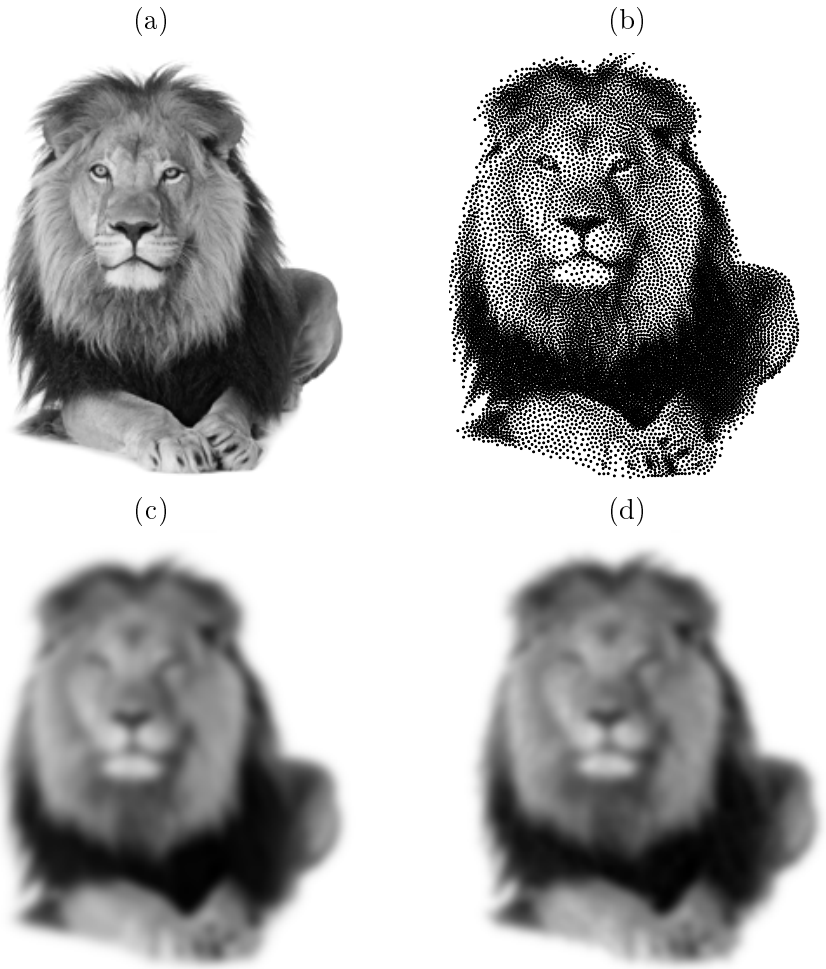


FIGURE 7.5: Illustration du problème de projection. Distribution $\pi$ (a) et représentation de la solution du problème (b). Représentation de la convolution par $h$.

$\{p_\star\gamma, p \in \mathcal{P}\}$ l'ensemble des mesures portées par les paramétrisations de $\mathcal{P}$, où $\gamma$ est la mesure de Lebesgue normalisée sur $[0,T]$. Un exemple de résultats théorique démontré dans ce papier est que si $P_T$ est un ensemble de courbes de $[0,T]$ dans $\Omega$ avec des contraintes sur ses dérivées en norme infinie (e.g., $\sup_t \|p^{(k)}\|_\infty \leqslant \alpha_k, 0 \leqslant k \leqslant N$), alors si $\mu_T$ est la solution du problème (7.2) pour $\mathcal{M}_N = \mathcal{M}(\mathcal{P}_T)$, $\mu_T$ converge faiblement vers $\pi$ quand $T \to \infty$.

Une application possible de ce résultat est la représentation d'images par une ligne continue (Fig. 7.6).

FIGURE 7.6: Exemple de projection sur un ensemble de mesures portées par des courbes.

Après discrétisation, le critère à minimiser s'écrit comme différence de deux fonctions convexes:

$$\min_{p \in P_N} \underbrace{\frac{1}{2N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} H(p_i - p_j)}_{\text{potentiel de répulsion}} - \underbrace{\frac{1}{N} \sum_{i=1}^{N} \int_{\Omega} H(x - p_i)\, d\pi(x)}_{\text{potentiel d'attraction}},$$

où $P_N$ est une discrétisation de $\mathcal{P}_T$ par $N$ points et $H$ est un noyau relié à $h$ par $\hat{H}(\xi) = |\hat{h}|^2(\xi)$. Un exemple de noyau possible est $H(x) = -\|x\|_2$. En pratique, nous utilisons une norme $\ell_2$ avec une régularisation en 0 pour la rendre infiniment dérivable. Ainsi les deux termes sont différentiables, et si nous supposons que $P_N$ est convexe, un algorithme de descente de gradient projeté converge vers un point critique de la fonction à minimiser. L'étape de projection sur l'ensemble des contraintes se fait en utilisant l'algorithme présenté au chapitre précédent.

## Résumé du chapitre 6

L'objectif de ce chapitre est d'appliquer les idées de projection de mesures dans le cas de simulations pour l'IRM. Nous fixons une mesure empirique $p$ qui fournit des résultats de reconstruction satisfaisants dans le cadre de tirages i.i.d.. Nous effectuons ensuite des projections de la mesure $p$ sur différents espaces de mesures:

- Des sommes de mesures de Dirac (Fig. 7.7(a)). Les schémas obtenus sont composés de mesures isolées, et améliorent (légèrement) les résultats de reconstruction par rapport aux tirages i.i.d. ($\leqslant 1$ dB).
- Des mesures portées par des lignes de longueur variables (Fig. 7.7(b)). Ces trajectoires ne sont pas toujours physiquement plausibles, car les contraintes sur la vitesse ne sont pas prises en compte.
- Des mesures portées par des trajectoires de l'ensemble $\mathcal{S}_T$ (Fig. 7.7(c)). Cet ensemble est le plus grand ensemble contenant les trajectoires physiquement plausibles et exploite au maximum les capacités de la machine.

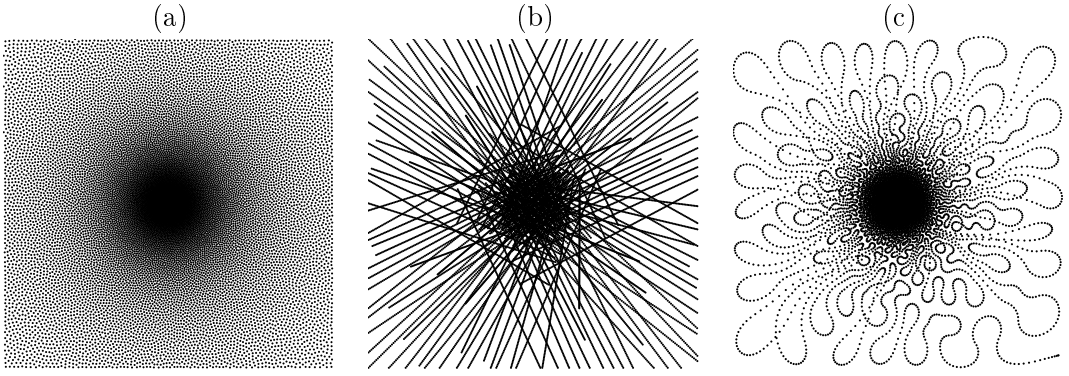


FIGURE 7.7: Exemple de schémas d'échantillonnage obtenus par projection de mesures sur : (a): une somme de mesures de Dirac; (b): des mesures portées par des lignes; (c): des mesures portées par des courbes admissibles de $\mathcal{S}_T$.

Les schéma continus obtenus par projection de mesures (Fig. 7.7(b-c)) permettent d'obtenir des résultats de reconstructions meilleurs que les stratégies classiques (spirales, radiales) avec un gain de l'ordre de 2 à 3 dB.

# Bibliography


Adcock, B., Hansen, A., Poon, C., and Roman, B. (2013). Breaking the coherence barrier: asymptotic incoherence and asymptotic sparsity in compressed sensing. *arXiv preprint arXiv:1302.0561*.

Adcock, B. and Hansen, A. C. (2011). Generalized sampling and infinite-dimensional compressed sensing. Technical report, Technical report NA2011/02, DAMTP, University of Cambridge.

Ahlswede, R. and Winter, A. (2002). Strong converse for identification via quantum channels. *Information Theory, IEEE Transactions on*, 48(3):569–579.

Aja-Fernández, S. and Tristán-Vega, A. (2013). A review on statistical noise models for magnetic resonance imaging. *LPI, ETSI Telecomunicacion, Universidad de Valladolid, Spain, Tech. Rep.*

Applegate, D., Bixby, R., Chvatal, V., and Cook, W. (2006). Concorde TSP solver. *URL: http://www.tsp.gatech.edu/concorde*.

Arias-Castro, E., Candes, E. J., Davenport, M., et al. (2013). On the fundamental limits of adaptive sensing. *Information Theory, IEEE Transactions on*, 59(1):472–481.

Attouch, H., Bolte, J., and Svaiter, B. F. (2013). Convergence of descent methods for semi-algebraic and tame problems: proximal algorithms, forward–backward splitting, and regularized Gauss–Seidel methods. *Mathematical Programming*, 137(1-2):91–129.

Aubert, G., Aujol, J.-F., and Blanc-Féraud, L. (2005). Detecting codimension-two objects in an image with Ginzburg-Landau models. *International Journal of Computer Vision*, 65(1-2):29–42.

Badillo, S., Vincent, T., and Ciuciu, P. (2013). Group-level impacts of within– and between–subject hemodynamic variability in fMRI. *Neuroimage*, 82:433–448.

Beardwood, J., Halton, J. H., and Hammersley, J. M. (1959). The shortest path through many points. In *Mathematical Proceedings of the Cambridge Philosophical Society*, volume 55, pages 299–327. Cambridge Univ Press.

Beck, A. and Teboulle, M. (2009a). A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM Journal on Imaging Sciences*, 2(1):183–202.

Beck, A. and Teboulle, M. (2009b). Gradient-based algorithms with applications to signal recovery. *Convex Optimization in Signal Processing and Communications*.

Beck, A. and Teboulle, M. (2014). A fast dual proximal gradient algorithm for convex minimization and applications. *Operations Research Letters*, 42(1):1–6.

Bigot, J., Boyer, C., and Weiss, P. (2013). An analysis of block sampling strategies in compressed sensing. *arXiv preprint arXiv:1305.4446*.

Billingsley, P. (2009). *Convergence of probability measures*, volume 493. Wiley.

Bloch, F. (1946). Nuclear induction. *Physical review*, 70(7-8):460.

Block, K. T., Uecker, M., and Frahm, J. (2007). Undersampled radial MRI with multiple coils. Iterative image reconstruction using a total variation constraint. *Magnetic Resonance in Medicine*, 57(6):1086–1098.

Bogachev, V. I. and Ruas, M. A. S. (2007). *Measure theory I*, volume 1. Springer.

Bosch, R. and Herman, A. (2004). Continuous line drawings via the traveling salesman problem. *Operations Research Letters*, 32(4):302–303.

Boyer, C., Bigot, J., and Weiss, P. (2015a). Compressed sensing with structured sparsity and structured acquisition. *arXiv preprint arXiv:1505.01619*.

Boyer, C., Bigot, J., and Weiss, P. (2015b). Compressed sensing with structured sparsity and structured acquisition. *Preprint*.

Boyer, C., Ciuciu, P., Weiss, P., and Mériaux, S. (2012). HYR2PICS: hybrid regularized reconstruction for combined parallel imaging and compressive sensing in MRI. In *Proc. of 9th IEEE ISBI conference*, pages 66–69, Barcelona, Spain.

Boyer, C., Weiss, P., and Bigot, J. ((in press) 2014). An algorithm for variable density sampling with block-constrained acquisition. *SIAM Journal on Imaging Science*.

Brémaud, P. (1999). *Markov chains: Gibbs fields, Monte Carlo simulation, and queues*, volume 31. springer.

Bridson, R. (2007). Fast Poisson disk sampling in arbitrary dimensions. In *ACM SIGGRAPH*, volume 2007, page 5.

Brodsky, E. K., Samsonov, A. A., and Block, W. F. (2009). Characterizing and correcting gradient errors in non-Cartesian imaging: Are gradient errors linear time-invariant (LTI)? *Magnetic Resonance in Medicine*, 62(6):1466–1476.

Brown, R. W., Cheng, Y.-C. N., Haacke, E. M., Thompson, M. R., and Venkatesan, R. (2014). *Magnetic resonance imaging: physical principles and sequence design*. John Wiley & Sons.

Candès, E., Romberg, J., and Tao, T. (2006a). Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. *Information Theory, IEEE Transactions on*, 52(2):489–509.

Candès, E., Romberg, J., and Tao, T. (2006b). Stable signal recovery from incomplete and inaccurate measurements. *Communications on pure and applied mathematics*, 59(8):1207–1223.

Candès, E. and Tao, T. (2006). Near optimal signal recovery from random projections: universal encoding strategies. *Information Theory, IEEE Transactions on*, 52(12):5406–5425.

Candès, E. J. (2008). The restricted isometry property and its implications for compressed sensing. *Comptes Rendus de l'Acad. des Sci.*, Serie I(346):589–592.

Candès, E. J. and Plan, Y. (2011). A probabilistic and RIPless theory of compressed sensing. *Information Theory, IEEE Transactions on*, 57(11):7235–7254.

Chang, C.-H. and Ji, J. (2010). Compressed sensing MRI with multichannel data using multicore processors. *Magnetic Resonance in Medicine*, 64(4):1135–1139.

Chauffert, N., Ciuciu, P., Kahn, J., and Weiss, P. (2013a). Travelling salesman-based compressive sampling. In *Proc. of 10th SampTA conference*, pages 509–511, Bremen, Germany.

Chauffert, N., Ciuciu, P., Kahn, J., and Weiss, P. (2014a). Variable density sampling with continuous trajectories. Application to MRI. *SIAM Journal on Imaging Science*, 7(4):1962–1992.

Chauffert, N., Ciuciu, P., Kahn, J., and Weiss, P. (2015a). A projection method on measures sets. *Preprint*.

Chauffert, N., Ciuciu, P., and Weiss, P. (2013b). Variable density compressed sensing in MRI. Theoretical vs. heuristic sampling strategies. In *Proc. of 10th IEEE ISBI conference*, pages 298–301, San Francisco, USA.

Chauffert, N., Ciuciu, P., Weiss, P., and Gamboa, F. (2013c). From variable density sampling to continuous sampling using Markov chains. In *Proc. of 10th SampTA conference*, pages 200–203, Bremen, Germany.

Chauffert, N., Weiss, P., Boucher, M., Mériaux, S., and Ciuciu, P. (2015b). Variable density sampling based on physically plausible gradient waveform. Application to 3D MRI angiography. In *IEEE International Symposium on Biomedical Imaging (ISBI)*.

Chauffert, N., Weiss, P., Kahn, J., and Ciuciu, P. (2014b). A projection algorithm for gradient waveforms design in magnetic resonance imaging. *Technical report.* *http://chauffertn.free.fr/Publis/2014/GradientWaveformDesign.pdf*.

Ciuciu, P., Poline, J.-B., Marrelec, G., Idier, J., Pallier, C., and Benali, H. (2003). Unsupervised robust nonparametric estimation of the hemodynamic response function for any fMRI experiment. *Medical Imaging, IEEE Transactions on*, 22(10):1235–1251.

Combettes, P. L. and Pesquet, J.-C. (2011). Proximal splitting methods in signal processing. In *Fixed-Point Algorithms for Inverse Problems in Science and Engineering*, pages 185–212. Springer.

Curtis, A. T. and Anand, C. K. (2008). Random volumetric MRI trajectories via genetic algorithms. *Journal of Biomedical Imaging*, 2008:6.

Daugman, J. G. (1980). Two-dimensional spectral analysis of cortical receptive field profiles. *Vision research*, 20(10):847–856.

Davids, M., Ruttorf, M., Zoellner, F., and Schad, L. (2015). Fast and robust design of time-optimal $k$-space trajectories in MRI. *Medical Imaging, IEEE Transactions on*, 34:564–577.

Diaconis, P. and Stroock, D. (1991). Geometric bounds for eigenvalues of Markov chains. *The Annals of Applied Probability*, 1(1):36–61.

Dinwoodie, I. H. (1995). A probability inequality for the occupation measure of a reversible Markov chain. *The Annals of Applied Probability*, 5(1):37–43.

Donoho, D. L. (2006). Compressed sensing. *Information Theory, IEEE Transactions on*, 52(4):1289–1306.

Duval, V. and Peyré, G. (2013). Exact support recovery for sparse spikes deconvolution. *arXiv preprint arXiv:1306.6909*.

Fatterpekar, G. M., Naidich, T. P., Delman, B. N., Aguinaldo, J. G., Gultekin, S. H., Sherwood, C. C., Hof, P. R., Drayer, B. P., and Fayad, Z. A. (2002). Cytoarchitecture

of the human cerebral cortex: MR microscopy of excised specimens at 9.4 Tesla. *American Journal of Neuroradiology*, 23(8):1313–1321.

Feinberg, D. A., Moeller, S., Smith, S. M., Auerbach, E., Ramanna, S., Gunther, M., Glasser, M. F., Miller, K. L., Ugurbil, K., and Yacoub, E. (2010). Multiplexed echo planar imaging for sub-second whole brain fMRI and fast diffusion imaging. *PloS one*, 5(12):e15710.

Feng, L., Grimm, R., Block, K. T., Chandarana, H., Kim, S., Xu, J., Axel, L., Sodickson, D. K., and Otazo, R. (2014). Golden-angle radial sparse parallel MRI: Combination of compressed sensing, parallel imaging, and golden-angle radial sampling for fast and flexible dynamic volumetric MRI. *Magnetic resonance in medicine*, 72(3):707–717.

Florescu, A., Chouzenoux, E., Pesquet, J.-C., Ciuciu, P., and Ciochina, S. (2014). A majorize-minimize memory gradient method for complex-valued inverse problems. *Signal Processing*, 103:285–295.

Fornasier, M., Haškovec, J., and Steidl, G. (2013). Consistency of variational continuous-domain quantization via kinetic theory. *Applicable Analysis*, 92(6):1283–1298.

Fornasier, M. and Hütter, J.-C. (2013). Consistency of probability measure quantization by means of power repulsion-attraction potentials. *arXiv preprint arXiv:1310.1120*.

Foucart, S. and Rauhut, H. (2013). A mathematical introduction to compressive sensing. *Appl. Numer. Harmon. Anal. Birkhäuser, Boston*.

Frangi, A. F., Niessen, W. J., Vincken, K. L., and Viergever, M. A. (1998). Multiscale vessel enhancement filtering. In *Medical Image Computing and Computer-Assisted Interventation (MICCAI 98)*, pages 130–137. Springer.

Freiberger, M., Knoll, F., Bredies, K., Scharfetter, H., and Stollberger, R. (2013). The agile library for biomedical image reconstruction using GPU acceleration. *Computing in Science & Engineering*, 15(1):34–44.

Frieze, A. M. and Yukich, J. E. (2002). Probabilistic analysis of the TSP. In Gutin, G. and Punnen, A. P., editors, *The traveling salesman problem and its variations*, volume 12 of *Combinatorial optimization*, pages 257–308. Springer.

Gillman, D. (1993). *Hidden Markov chains: Convergence rates and the complexity of inference*. PhD thesis, MIT.

Golden, S. (1965). Lower bounds for the Helmholtz function. *Physical Review*, 137(4B):B1127.

Gräf, M., Potts, D., and Steidl, G. (2012). Quadrature errors, discrepancies, and their relations to halftoning on the torus and the sphere. *SIAM Journal on Scientific Computing*, 34(5):A2760–A2791.

Griswold, M. A., Jakob, P. M., Heidemann, R. M., Nittka, M., Jellus, V., Wang, J., Kiefer, B., and Haase, A. (2002). Generalized autocalibrating partially parallel acquisitions GRAPPA. *Magnetic Resonance in Medicine*, 47(6):1202–1210.

Gröchenig, K., Romero, J. L., Unnikrishnan, J., and Vetterli, M. (2014). On minimal trajectories for mobile sampling of bandlimited fields. *Applied and Computational Harmonic Analysis*.

Gross, D. (2011). Recovering low-rank matrices from few coefficients in any basis. *Information Theory, IEEE Transactions on*, 57(3):1548–1566.

Gruber, P. M. (2004). Optimum quantization and its applications. *Advances in Mathematics*, 186(2):456–497.

Guerquin-Kern, M., Häberlin, M., Pruessmann, K. P., and Unser, M. (2011). A fast wavelet-based reconstruction method for magnetic resonance imaging. *Medical Imaging, IEEE Transactions on*, 30(9):1649–1660.

Guerquin-Kern, M., Lejeune, L., Pruessmann, K. P., and Unser, M. (2012). Realistic analytical phantoms for parallel magnetic resonance imaging. *Medical Imaging, IEEE Transactions on*, 31(3):626–636.

Gurney, P. T., Hargreaves, B. A., and Nishimura, D. G. (2006). Design and analysis of a practical 3D cones trajectory. *Magn. Reson. Med.*, 55(3):575–582.

Gwosdek, P., Schmaltz, C., Weickert, J., and Teuber, T. (2014). Fast electrostatic halftoning. *Journal of real-time image processing*, 9(2):379–392.

Haldar, J. P., Hernando, D., and Liang, Z.-P. (2011). Compressed-sensing MRI with random encoding. *Medical Imaging, IEEE Transactions on*, 30(4):893–903.

Handwerker, D. A., Ollinger, J. M., and D'Esposito, M. (2004). Variation of BOLD hemodynamic responses across subjects and brain regions and their effects on statistical analyses. *Neuroimage*, 21(4):1639–1651.

Hargreaves, B. A., Nishimura, D. G., and Conolly, S. M. (2004). Time-optimal multidimensional gradient waveform design for rapid imaging. *Magnetic Resonance in Medicine*, 51(1):81–92.

Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57(1):97–109.

Hiriart-Urruty, J.-B. (2009). A new series of conjectures and open questions in optimization and matrix analysis. *ESAIM: Control, Optimisation and Calculus of Variations*, 15(02):454–470.

Hiriart-Urruty, J.-B. and Lemaréchal, C. (1996). *Convex Analysis and Minimization Algorithms I: Part 1: Fundamentals*, volume 305. Springer.

Horn, R. and Johnson, C. (1991). *Topics in matrix analysis*. Cambridge University Press, Cambridge.

Huang, Y., Paisley, J., Lin, Q., Ding, X., Fu, X., and Zhang, X.-P. (2014). Bayesian nonparametric dictionary learning for compressed sensing MRI. *Image Processing, IEEE Transactions on*, 23(12):5007–5019.

Jackson, J. I., Meyer, C. H., Nishimura, D. G., and Macovski, A. (1991). Selection of a convolution function for Fourier inversion using gridding [computerised tomography application]. *Medical Imaging, IEEE Transactions on*, 10(3):473–478.

Jerrum, M. and Sinclair, A. (1989). Approximating the permanent. *SIAM journal on computing*, 18(6):1149–1178.

Joulin, A., Ollivier, Y., et al. (2010). Curvature, concentration and error estimates for Markov chain Monte Carlo. *The Annals of Probability*, 38(6):2418–2442.

Juditsky, A., Karzan, F., and Nemirovski, A. (2011). On low rank matrix approximations with applications to synthesis problem in compressed sensing. *SIAM J. on Matrix Analysis and Applications*, 32(3):1019–1029.

Juditsky, A. and Nemirovski, A. (2011). On verifiable sufficient conditions for sparse signal recovery via $\ell_1$ minimization. *Mathematical Programming Ser. B*, 127:89–122.

Kaplan, C. S., Bosch, R., et al. (2005). Tsp art. In *Renaissance Banff: Mathematics, Music, Art, Culture*, pages 301–308. Canadian Mathematical Society.

Kargin, V. (2007). A large deviation inequality for vector functions on finite reversible Markov chains. *The Annals of Applied Probability*, pages 1202–1221.

Kato, T. (1976). *Perturbation theory for linear operators*, volume 132. Springer Science & Business Media.

Katznelson, Y. (1968). An introduction to harmonic analysis. *New York*.

Keiner, J., Kunis, S., and Potts, D. (2009). Using NFFT 3—a software library for various nonequispaced fast Fourier transforms. *ACM Transactions on Mathematical Software (TOMS)*, 36(4):19.

Kim, D. H., Adalsteinsson, E., and Spielman, D. M. (2003). Simple analytic variable density spiral design. *Magnetic Resonance in Medicine*, 50(1):214–219.

Kloeckner, B. (2012). Approximation by finitely supported measures. *ESAIM: Control, Optimisation and Calculus of Variations*, 18(02):343–359.

Knoll, F., Clason, C., Diwoky, C., and Stollberger, R. (2011). Adapted random sampling patterns for accelerated MRI. *Magma*, 24(1):43–50.

Knopp, T., Kunis, S., and Potts, D. (2007). A note on the iterative MRI reconstruction from nonuniform $k$-space data. *International journal of biomedical imaging*, 2007.

Krahmer, F. and Ward, R. (2014). Stable and robust sampling strategies for compressive imaging. *Image Processing, IEEE Transactions on*, 23(2):612–622.

Kumar Anand, C., Thomas Curtis, A., and Kumar, R. (2008). Durga: A heuristically-optimized data collection strategy for volumetric magnetic resonance imaging. *Engineering Optimization*, 40(2):117–136.

Kurdyka, K. (1998). On gradients of functions definable in o-minimal structures. In *Annales de l'institut Fourier*, volume 48, pages 769–783. Institut Fourier.

Lauterbur, P. C. et al. (1973). Image formation by induced local interactions: examples employing nuclear magnetic resonance. *Nature*, 242(5394):190–191.

Le Ny, J., Feron, E., and Frazzoli, E. (2012). On the Dubins traveling salesman problem. *Automatic Control, IEEE Transactions on*, 57(1):265–270.

Lezaud, P. (1998). Chernoff-type bound for finite Markov chains. *Annals of Applied Probability*, pages 849–867.

Li, H. and Mould, D. (2014). Continuous line drawings and designs. *International Journal of Creative Interfaces and Computer Graphics*.

Liang, Z.-P. and Lauterbur, P. C. (2000). *Principles of magnetic resonance imaging*. SPIE Optical Engineering Press.

Lieb, E. H. (1973). Convex trace functions and the Wigner-Yanase-Dyson conjecture. *Advances in Mathematics*, 11(3):267–288.

Liu, D.-D., Liang, D., Liu, X., and Zhang, Y.-T. (2012). Under-sampling trajectory design for compressed sensing MRI. In *Engineering in Medicine and Biology Society (EMBC), 2012 Annual International Conference of the IEEE*, pages 73–76.

Lustig, M., Donoho, D. L., and Pauly, J. M. (2007). Sparse MRI: The application of compressed sensing for rapid MR imaging. *Magnetic Resonance in Medicine*, 58(6):1182–1195.

Lustig, M., Kim, S. J., and Pauly, J. M. (2008). A fast method for designing time-optimal gradient waveforms for arbitrary $k$-space trajectories. *Medical Imaging, IEEE Transactions on*, 27(6):866–873.

Lustig, M., Lee, J. H., Donoho, D. L., and Pauly, J. M. (2005). Faster imaging with randomly perturbed, under-sampled spirals and $\ell_1$ reconstruction. In *Proceedings of the 13th annual meeting of ISMRM, Miami Beach*, page 685.

Ma, S., Yin, W., Zhang, Y., and Chakraborty, A. (2008). An efficient algorithm for compressed MR imaging using total variation and wavelets. In *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, pages 1–8. IEEE.

Mackey, L., Jordan, M. I., Chen, R. Y., Farrell, B., Tropp, J. A., et al. (2014). Matrix concentration inequalities via the method of exchangeable pairs. *The Annals of Probability*, 42(3):906–945.

Majewski, K., Heid, O., and Kluge, T. (2010). MRI pulse sequence design with first-order gradient moment nulling in arbitrary directions by solving a polynomial program. *IEEE Trans. Med. Imag.*, 29(6):1252–1259.

Mallat, S. (1999). *A wavelet tour of signal processing*. Academic press.

Marim, M. M., Atlan, M., Angelini, E., and Olivo-Marin, J.-C. (2010). Compressed sensing with off-axis frequency-shifting holography. *Optics letters*, 35(6):871–873.

Marteniuk, R., MacKenzie, C., Jeannerod, M., Athenes, S., and Dugas, C. (1987). Constraints on human arm movement trajectories. *Canadian Journal of Psychology/Revue canadienne de psychologie*, 41(3):365.

Maudsley, A. (1988). Dynamic range improvement in NMR imaging using phase scrambling. *Journal of Magnetic Resonance (1969)*, 76(2):287–305.

Mir, R., Guesalaga, A., Spiniak, J., Guarini, M., and Irarrazaval, P. (2004). Fast three-dimensional $k$-space trajectory design using missile guidance ideas. *Magnetic Resonance in Medicine*, 52(2):329–336.

Mordukhovich, B. S. (2006). *Variational Analysis and Generalized Differentiation I: Basic Theory*, volume 330. Springer.

Murphy, M., Alley, M., Demmel, J., Keutzer, K., Vasanawala, S., and Lustig, M. (2012). Fast-SPIRiT compressed sensing parallel imaging MRI: scalable parallel implementation and clinically feasible runtime. *Medical Imaging, IEEE Transactions on*, 31(6):1250–1262.

Nesterov, Y. (1983). A method of solving a convex programming problem with convergence rate $\mathcal{O}(1/k^2)$. In *Soviet Mathematics Doklady*, volume 27, pages 372–376.

Nesterov, Y. (2005). Smooth minimization of non-smooth functions. *Math. Program.*, 103(1):127–152.

Nesterov, Y. (2013). Gradient methods for minimizing composite functions. *Mathematical Programming*, 140(1):125–161.

Nishimura, D. G., Irarrazabal, P., and Meyer, C. H. (1995). A velocity $k$-space analysis of flow effects in echo-planar and spiral imaging. *Magnetic Resonance in Medicine*, 33(4):549–556.

Noll, D. C. (1997). Multishot rosette trajectories for spectrally selective MR imaging. *Medical Imaging, IEEE Transactions on*, 16(4):372–377.

Oliveira, R. I. (2009). Concentration of the adjacency matrix and of the Laplacian in random graphs with independent edges. *arXiv preprint arXiv:0911.0600*.

Oliveira, R. I. (2010). Sums of random Hermitian matrices and an inequality by Rudelson. *Electron. Commun. Probab*, 15(203-212):26.

O'sullivan, J. (1985). A fast sinc function gridding algorithm for Fourier inversion in computer tomography. *Medical Imaging, IEEE Transactions on*, 4(4):200–207.

Pappas, T. N. and Neuhoff, D. L. (1999). Least-squares model-based halftoning. *Image Processing, IEEE Transactions on*, 8(8):1102–1116.

Park, J., Zhang, Q., Jellus, V., Simonetti, O., and Li, D. (2005). Artifact and noise suppression in GRAPPA imaging using improved $k$-space coil calibration and variable density sampling. *Magnetic Resonance in Medicine*, 53(1):186–193.

Paulin, D. (2012a). Concentration inequalities for Markov chains by Marton couplings. *arXiv preprint arXiv:1212.2015*.

Paulin, D. (2012b). A note on matrix concentration inequalities via the method of exchangeable pairs. *arXiv preprint arXiv:1212.2012*.

Paulin, D. (2014). Mixing and concentration by Ricci curvature. *arXiv preprint arXiv:1404.2802*.

Paulin, D., Mackey, L., and Tropp, J. A. (2013). Deriving matrix concentration inequalities from kernel couplings. *arXiv preprint arXiv:1305.0612*.

Pipe, J. G. and Zwart, N. R. (2014). Spiral trajectory design: A flexible numerical algorithm and base analytical equations. *Magn. Reson. Med.*, 71(1):278–285.

Polak, A. C., Duarte, M. F., and Goeckel, D. L. (2012a). Grouped incoherent measurements for compressive sensing. In *Statistical Signal Processing Workshop (SSP), 2012 IEEE*, pages 732–735. IEEE.

Polak, A. C., Duarte, M. F., and Goeckel, D. L. (2012b). Performance bounds for grouped incoherent measurements in compressive sensing. *arXiv preprint arXiv:1205.2118*.

Poole, M. S. (2007). *Improved equipment and techniques for dynamic shimming in high field MRI*. PhD thesis, University of Nottingham.

Poon, C. (2015a). On the role of total variation in compressed sensing. *SIAM Journal on Imaging Sciences*, 8(1):682–720.

Poon, C. (2015b). Structure dependent sampling in compressed sensing: theoretical guarantees for tight frames. *arXiv preprint arXiv:1505.05291*.

Potts, D. and Steidl, G. (2003). Fast summation at nonequispaced knots by NFFT. *SIAM Journal on Scientific Computing*, 24(6):2013–2037.

Pruessmann, K. P., Weiger, M., Scheidegger, M. B., Boesiger, P., et al. (1999). SENSE: sensitivity encoding for fast MRI. *Magnetic Resonance in Medicine*, 42(5):952–962.

Purcell, E. M., Torrey, H., and Pound, R. V. (1946). Resonance absorption by nuclear magnetic moments in a solid. *Physical review*, 69(1-2):37.

Puy, G., Marques, J. P., Gruetter, R., Thiran, J., Van De Ville, D., Vandergheynst, P., and Wiaux, Y. (2012a). Spread spectrum magnetic resonance imaging. *Medical Imaging, IEEE Transactions on*, 31(3):586–598.

Puy, G., Vandergheynst, P., Gribonval, R., and Wiaux, Y. (2012b). Universal and efficient compressed sensing by spread spectrum and application to realistic Fourier imaging techniques. *EURASIP Journal on Advances in Signal Processing*, 2012(1):1–13.

Puy, G., Vandergheynst, P., and Wiaux, Y. (2011). On variable density compressive sampling. *Signal Processing letters, IEEE Transactions on*, 18(10):595–598.

Rauhut, H. (2010). Compressive Sensing and Structured Random Matrices. In Fornasier, M., editor, *Theoretical Foundations and Numerical Methods for Sparse Recovery*, volume 9 of *Radon Series Comp. Appl. Math.*, pages 1–92. deGruyter.

Ravishankar, S. and Bresler, Y. (2011a). Adaptive sampling design for compressed sensing MRI. In *Engineering in Medicine and Biology Society, EMBC, 2011 Annual International Conference of the IEEE*, pages 3751–3755.

Ravishankar, S. and Bresler, Y. (2011b). MR image reconstruction from highly undersampled $k$-space data by dictionary learning. *Medical Imaging, IEEE Transactions on*, 30(5):1028–1041.

Rivenson, Y., Stern, A., and Javidi, B. (2010). Compressive Fresnel holography. *Display Technology, Journal of*, 6(10):506–509.

Rockafellar, R. T. (1997). *Convex analysis*. Number 28. Princeton university press.

Roman, B., Hansen, A., and Adcock, B. (2014). On asymptotic structure in compressed sensing. *arXiv preprint arXiv:1406.4178*.

Romberg, J. (2009). Compressive sensing by random convolution. *SIAM Journal on Imaging Sciences*, 2(4):1098–1128.

Sarty, G. E., Bennett, R., and Cox, R. W. (2001). Direct reconstruction of non-Cartesian $k$-space data using a nonuniform fast Fourier transform. *Magnetic Resonance in Medicine*, 45(5):908–915.

Schmaltz, C., Gwosdek, P., Bruhn, A., and Weickert, J. (2010). Electrostatic halftoning. In *Computer Graphics Forum*, volume 29, pages 2313–2327. Wiley Online Library.

Seeger, M., Nickisch, H., Pohmann, R., and Schölkopf, B. (2010). Optimization of $k$-space trajectories for compressed sensing by Bayesian experimental design. *Magnetic Resonance in Medicine*, 63(1):116–126.

Shu, Y., Riederer, S. J., and Bernstein, M. A. (2006). Three-dimensional MRI with an undersampled spherical shells trajectory. *Magnetic resonance in medicine*, 56(3):553–562.

Sidky, E. Y., Kao, C.-M., and Pan, X. (2006). Accurate image reconstruction from few-views and limited-angle data in divergent-beam CT. *Journal of X-ray Science and Technology*, 14(2):119–139.

Simonetti, O. P., Duerk, J. L., and Chankong, V. (1993). An optimal design method for magnetic resonance imaging gradient waveforms. *IEEE Trans. Med. Imag.*, 12(2):350–360.

Smale, S. (1998). Mathematical problems for the next century. *The Mathematical Intelligencer*, 20(2):7–15.

Smith, D. S., Gore, J. C., Yankeelov, T. E., and Welch, E. B. (2012). Real-time compressive sensing MRI reconstruction using GPU computing and split Bregman methods. *International journal of biomedical imaging*.

Sodickson, D. K. and Manning, W. J. (1997). Simultaneous acquisition of spatial harmonics (SMASH): fast imaging with radiofrequency coil arrays. *Magnetic Resonance in Medicine*, 38(4):591–603.

Spielman, D. M., Pauly, J. M., and Meyer, C. H. (1995). Magnetic resonance fluoroscopy using spirals with variable sampling densities. *Magnetic Resonance in Medicine*, 34(3):388–394.

Spiniak, J., Guesalaga, A., Mir, R., Guarini, M., and Irarrazaval, P. (2005). Undersampling $k$-space using fast progressive 3D trajectories. *Magnetic Resonance in Medicine*, 54(4):886–892.

Stanisz, G. J., Odrobina, E. E., Pun, J., Escaravage, M., Graham, S. J., Bronskill, M. J., and Henkelman, R. M. (2005). $T_1$, $T_2$ relaxation and magnetization transfer in tissue at 3T. *Magnetic Resonance in Medicine*, 54(3):507–512.

Steele, J. M. (1981). Subadditive euclidean functionals and nonlinear growth in geometric probability. *The Annals of Probability*, 9(3):365–376.

Strang, G. and Nguyen, T. (1996). *Wavelets and filter banks*. SIAM.

Teuber, T., Steidl, G., Gwosdek, P., Schmaltz, C., and Weickert, J. (2011). Dithering by differences of convex functions. *SIAM Journal on Imaging Sciences*, 4(1):79–108.

Thompson, C. J. (1965). Inequality with applications in statistical mechanics. *Journal of Mathematical Physics*, 6(11):1812–1813.

Thomson, J. J. (1904). On the structure of the atom. *Philos. Mag., Ser. 6*, 7:237–265.

Tropp, J. A. (2012). User-friendly tail bounds for sums of random matrices. *Foundations of Computational Mathematics*, 12(4):389–434.

Tsai, C. M. and Nishimura, D. G. (2000). Reduced aliasing artifacts using variable-density $k$-space sampling trajectories. *Magnetic Resonance in Medicine*, 43(3):452–458.

Twieg, D. B. (1983). The $k$-trajectory formulation of the NMR imaging process with applications in analysis and synthesis of imaging methods. *Medical Physics*, 10(5):610–621.

Ulichney, R. (1987). *Digital halftoning*. MIT press.

Unnikrishnan, J. and Vetterli, M. (2013). Sampling high-dimensional bandlimited fields on low-dimensional manifolds. *Information Theory, IEEE Transactions on*, 59(4):2103–2127.

Vasanawala, S., Murphy, M., Alley, M. T., Lai, P., Keutzer, K., Pauly, J. M., and Lustig, M. (2011). Practical parallel imaging compressed sensing MRI: Summary of two years of experience in accelerating body MRI of pediatric patients. In *Proc. ISBI*, pages 1039–1043.

Vershynin, R. (2009). A note on sums of independent random matrices after Ahlswede-Winter. *Lecture notes*.

Villani, C. (2008). *Optimal transport: old and new*, volume 338. Springer.

Wang, H., Wang, X., Zhou, Y., Chang, Y., and Wang, Y. (2012). Smoothed random-like trajectory for compressed sensing MRI. In *Proc. of the 34th annual IEEE EMBC*, pages 404–407.

Weiss, P., Blanc-Féraud, L., and Aubert, G. (2009). Efficient schemes for total variation minimization under constraints in image processing. *SIAM journal on Scientific Computing*, 31(3):2047–2080.

Wiaux, Y., Puy, G., Boursier, Y., and Vandergheynst, P. (2009). Spread spectrum for imaging techniques in radio interferometry. *Monthly Notices of the Royal Astronomical Society*, 400(2):1029–1038.

Willett., R. M. (2011). Errata: Sampling trajectories for sparse image recovery. Note, Duke University.

Willett, R. M. (2011). Short and smooth sampling trajectories for compressed sensing. In *Acoustics, Speech and Signal Processing (ICASSP), 2011 IEEE International Conference on*, pages 4012–4015. IEEE.

Winkelmann, S., Schaeffter, T., Koehler, T., Eggers, H., and Doessel, O. (2007). An optimal radial profile order based on the golden ratio for time-resolved MRI. *Medical Imaging, IEEE Transactions on*, 26(1):68–76.

Wong, F. J. and Takahashi, S. (2011). A graph-based approach to continuous line illustrations with variable levels of detail. In *Computer Graphics Forum*, volume 30, pages 1931–1939. Wiley Online Library.

Xu, J. and Kaplan, C. S. (2007). Image-guided maze construction. In *ACM Transactions on Graphics (TOG)*, volume 26, page 29. ACM.

Yukich, J. E. (1998). *Probability theory of classical Euclidean optimization problems*. Springer.